



UNIVERSITAT DE
BARCELONA

Facultat de Matemàtiques
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

Formalització i modelització d'una
comunitat virtual enriquida amb
un xatbot

Autor: Iago Pueyo Puyuelo

Director: Dra. Maite López Sánchez

Director: Dr. Jerónimo Hernández González

Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 24 de gener de 2023

Abstract

Collaboration is increasingly occurring in virtual online communities. In them, the forms of interaction are varied, each of the virtual communities has its own way of supporting the activities it proposes, the roles the participants take, the actions they can carry out, etc. This work formalizes the generic model that characterizes virtual communities and proposes to include a conversational agent (or chatbot) within the virtual community, with the aim of being able to understand the environment and generate interaction within it.

In addition, it is proposed to use a computational method based on probabilistic graphical models and machine learning. Through the inference of data corresponding to the visible interactions of the users, the variables of the model - which also includes variables that model non-visible interactions - would take on values, which would serve to predict behaviours. In particular, the virtual community of the platform `decidim.barcelona` is considered, which revolves around citizen participation, that is, the active intervention of citizens in public decision-making processes. The interaction spaces are analyzed and, with the aim of encouraging interaction, the incorporation of a chatbot with the ability to consult the model is proposed.

Resum

La col·laboració es produeix cada cop més en comunitats virtuals en línia. En elles, les formes d'interacció són variades, cadascuna de les comunitats virtuals té la seva forma de suportar les activitats que proposa, els rols que prenen les persones participants, les accions que puguin dur a terme, etc. Aquest treball formalitza el model genèric que caracteritza les comunitats virtuals i es proposa incloure un agent conversacional (o xatbot) dins la comunitat virtual, amb el propòsit que pugui entendre l'entorn i ser capaç de generar interacció dins d'aquest.

A més, es proposa utilitzar un mètode computacional basat en models gràfics probabilístics i aprenentatge automàtic. Mitjançant la inferència de dades corresponents a les interaccions visibles de les persones usuàries, les variables del model -que també inclou variables que modelen interaccions no visibles - prendrien valors, els quals servirien per poder predir comportaments. En particular, es considera la comunitat virtual de la plataforma `decidim.barcelona`, que gira entorn la participació ciutadana, és a dir, la intervenció activa de la ciutadania en els processos de presa de decisions públiques. S'analitzen els espais d'interacció i, amb l'objectiu de fomentar la interacció, es proposa la incorporació d'un xatbot amb la capacitat de consultar el model.

Agraïments

Vull agrair la bona predisposició i comunicació que la doctora Maite López i el doctor Jerónimo Hernández, tutors oficials d'aquest treball de final de grau, m'han brindat durant els semestres. Menció especial per la doctora Inmaculada Rodríguez, qui ha estat una tutora més. Tant en el moment de decidir l'enfocament del treball, com en l'execució d'aquest la seva ajuda m'ha facilitat i alleugerit el camí.

Per altra banda, estic molt agraït amb el suport que rebut per part de la meva família, dels amics amb els quals ens hem recolzat mútuament i d'aquells amb qui he compartit reflexions i dubtes sobre el treball. M'agradaria fer una menció especial al Guillem Quingles, la Paula Pastó i l'Axel Gómez

Índex

1	Introducció	1
1.1	Comunitats Virtuals: definició i propietats	1
1.2	Modelització i visualització	3
1.3	Xatbots en comunitat virtuals	5
1.4	Objectius	6
1.5	Estructura de la memòria	6
2	Estat de l'art	7
3	Formalització d'una comunitat virtual	9
3.1	Proposta	9
3.2	Exemples	16
3.2.1	Comunitats Virtuals en l'educació	16
3.2.2	Comunitats Virtuals de participació ciutadana	22
4	Model de comunitat virtual decidim.barcelona	27
4.1	Models Gràfics Probabilístics	27
4.1.1	Exemple	28
4.1.2	Xarxes Bayesianes	29
4.2	Model de la comunitat virtual decidim.barcelona	34
4.2.1	Variables	34
4.2.2	Estructura i probabilitats	38
5	Conclusions i futur treball	43
5.1	Conclusions	43
5.1.1	Consecució d'objectius	44
5.2	Línies de futur treball	44

A Preliminars	48
A.1 Teoria de probabilitat	48
A.2 Teoria de grafs	51

Capítol 1

Introducció

El desenvolupament del coneixement col·lectiu humà sempre ha depès de les comunitats [28]. A mesura que aquestes han esdevingut més tecnològiques, en el sentit que l'ús de l'ordinador s'ha anat normalitzant, ha fet augmentar les relacions en línia i, en conseqüència, les maneres en què les comunitats poden formar-se i com els seus membres poden interactuar.

Així com sona de beneficiosa la idea d'incloure altres formes de comunicació a una comunitat, també podria arribar a comprometre la sostenibilitat de sí mateixa quan es restringeix al ciberespai [18]. El sistema tecnològic que suporti la comunitat haurà de garantir una qualitat, tant de servei com d'informació, per tal de millorar la percepció de l'usuari, alimentant la motivació i el compromís i, així, fomentant la participació.

Entre els nous enfocaments que s'han anat introduint per mantenir els índexs participatius alts destaca la implementació de xatbots. Consisteixen en sistemes intel·ligents que aprofiten les tecnologies del llenguatge natural per mantenir converses basades en text amb persones usuàries [17]. És un potencial model d'aplicació tecnològica per promoure la comunicació i l'aprenentatge interpersonal oferint diferents tipus d'informació i coneixement mitjançant mètodes interactius i interfícies fàcils d'operar. L'àmbit d'aplicació també és divers: des de consulta de dades obertes [13], fins involucrar-se en les interaccions socials de diverses comunitats virtuals d'educació [14] o culturals [25], entre d'altres.

1.1 Comunitats Virtuals: definició i propietats

En un sentit ample, una comunitat virtual és un grup de persones amb objectius comuns que interactuen entre elles en un entorn online. Existeixen definicions més conceptualitzades, com la que es dona a [10]: *“una comunitat virtual és la combinació d'un sistema tècnic, l'objectiu del qual és maximitzar la realització d'activitats, i d'un sistema social, en què l'objectiu és maximitzar la qualitat de vida dels persones usuàries del sistema”*. El sistema social conté moltes construccions socials complexes com ara objectius, fluxos de treball, estructures organitzatives i normes

socials. El sistema social està recolzat pel sistema tècnic garantint bona connectivitat, qualitat de la informació i persistència de les dades que la pròpia comunitat genera.

En qualsevol cas, les comunitats virtuals s'estructuren al voltant de la interacció entre els membres. Aquesta interacció és causada per la participació de les persones usuàries en les diverses activitats que s'ofereixen: normalment poden ser *activitats informatives*, que es basen en la recopilació, distribució i utilització d'informació en qualsevol format; *activitats discursives*, com pot ser compartir l'opinió o entaular converses amb altres persones participants; o *activitats participatives*, que poden ser tant *individuals*, per exemple formar part d'una votació o respondre qüestionaris, com *col·laboratives*, com pot ser una videoconferència o involucrar-se en projectes d'equip. La diversitat d'activitats afavorirà la participació i la interacció de les persones usuàries i, així també, la socialització i l'aprenentatge interpersonal [27].

Majoritàriament, les activitats que tenen lloc a les comunitats virtuals estan significativament influenciades per dues característiques principals: en primer lloc gairebé totes les interaccions entre participants estan mediades per text; en segon lloc, aquestes mateixes interaccions es donen en línia, és a dir, com l'enviament de missatges no requereix proximitat física ni, a vegades, immediatesa, les persones participants poden estar físicament separades i no sempre connectades. En conseqüència d'aquestes dues característiques, sovint el tipus d'activitats que es donen a la virtualitat són asíncrones, fent que sigui un factor diferenciador respecte les tradicionals. Tal com suggereix [26] en una anàlisi comparativa de les comunitats virtuals i presencials, és probable que les diferents comunitats tinguin una manera única de donar suport a les activitats de les persones participants. A més, part de les singularitats d'una comunitat determinada es poden veure en les normes i patrons particulars de comportament acceptable dins de la comunitat. És a dir, cada comunitat fa èmfasi en els seus patrons particulars d'interacció i estableix les seves pròpies normes i expectatives.

Resumidament, una sèrie de propietats que caracteritzen les comunitats virtuals són:

- Tenen el potencial de donar suport a una gran varietat d'activitats relacionades amb l'objectiu de la comunitat (aprendre en una comunitat educativa, fer ciència en una comunitat de ciència ciutadana, etc.). Aquestes activitats es poden donar síncrona o asíncronament.
- Oferir la possibilitat de revisar l'històric: permet posar-se al dia de decisions preses mentre la persona usuària estava desconnectada o, en el cas d'un nova persona participant a la comunitat, beneficiar-se del coneixement generat prèviament.
- La comunicació es multidireccional, és a dir, és un tipus de comunicació que va en totes les direccions, essent cada integrant de les activitats emissor i receptor a la vegada, utilitzant un mateix canal per poder transmetre els missatges, facilitant la comunicació tant d'1 a 1 com d'1 a molts, o entre grups.

Al llarg del projecte reconeixem i tindrem en compte la influència de les anteriors propietats, observant analogies entre els dos models de comunitats, com pot ser les comunitats educatives d'una classe en línia en comparació amb una de presencial. Prendrem, a més, la comunitat virtual de la plataforma `decidim.barcelona` i observarem quines propietats la caracteritzen.

1.2 Modelització i visualització

Les relacions que s'estableixen a les comunitats poden ser visualment representades mitjançant grafs, on els nodes serien les persones que conformen la comunitat i les arestes, la connexió entre elles. Aquestes arestes poden incloure un pes i un sentit, com ho és el graf resultant de considerar el nombre de citacions entre autors. En un cas més simple, podem considerar el graf no dirigit 1.1 per veure les citacions entre diferents publicacions.

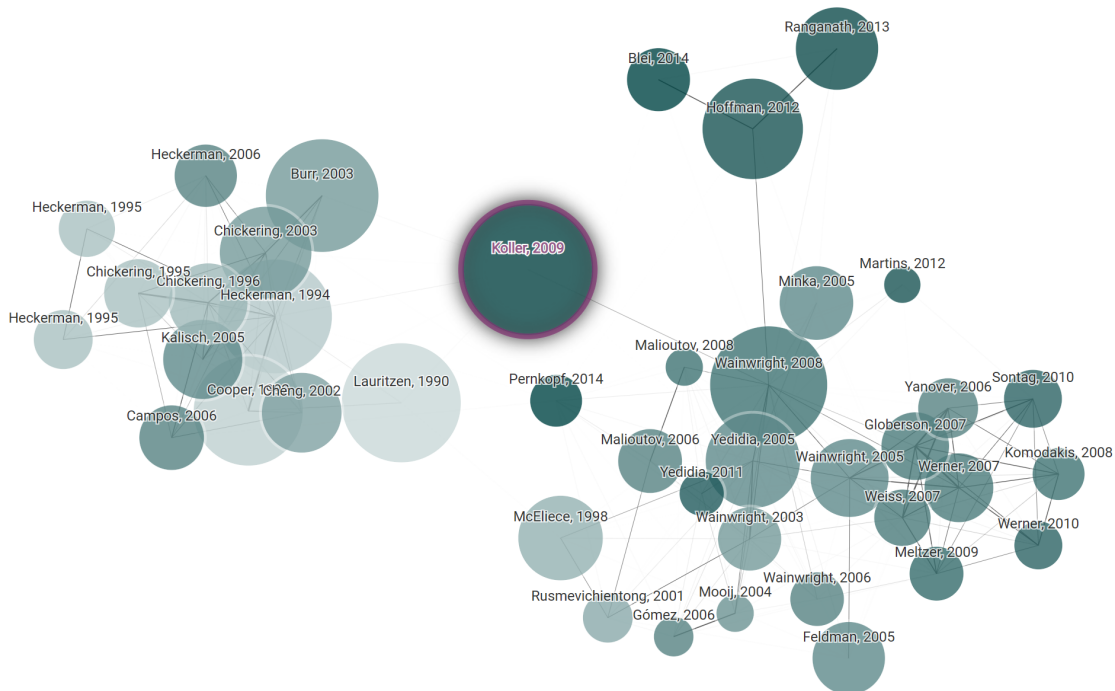


Figura 1.1: Graf resultant de considerar les citacions del llibre [16] i connexions entre aquestes. Ha estat obtingut amb l'eina www.connectedpapers.com

Més enllà de quedar-nos amb el concepte que la comunitat només està conformada pels seus membres i les relacions que s'estableixen, podem definir una sèrie de variables que influeixen en el comportament d'aquesta. La idea bàsica és identificar les variables i el tipus de relacions, tant d'independència com dependència, que són vàlides per a un determinat problema. Una manera natural de representar les relacions de dependència i independència entre un conjunt de variables és utilitzar grafs dirigits, de manera que les variables que són directament dependents

estan connectades i les relacions d'independència estan implícites en aquest gràfic de dependència.

Amb aquest plantejament, podem definir sobre les variables distribucions de probabilitat [16] que, alhora, depenen dels valors que prenen les variables que l'apunten, obtenint el que s'anomena model gràfic probabilístic (PGM, per les sigles en anglès). Un PGM és una representació compacta d'una distribució de probabilitat conjunta, predir comportaments d'una manera computacionalment eficient. Concretament, mitjançant l'observació d'algunes variables es podrà fer inferència sobre el model, és a dir, deduir els valors que prendran les variables que no es perceben.

Exemple 1.2.1. Suposem que un estudiant té una intel·ligència I amb domini $Val(I) = \{i^0, i^1\}$, i ha participat en un curs la dificultat del qual ve representada amb la variable aleatòria D amb domini $Val(D) = \{d^0, d^1\}$. La nota de l'estudiant G depèn de la seva intel·ligència i de la dificultat del curs, i tindrà domini $Val(G) = \{g^1, g^2, g^3\}$. Suposem també que l'alumne ha demanat una carta de recomanació i que aquesta només depèn de la nota que ha obtingut al curs. La qualitat de la seva lletra és una variable aleatòria L , el domini de la qual és $Val(L) = \{l^0, l^1\}$. Finalment, suposarem que l'alumne pot haver fet l'examen *SAT* el qual valora les capacitats i competències personals. Aquesta serà una variable aleatòria S que prendrà valors $Val(S) = \{s^0, s^1\}$.

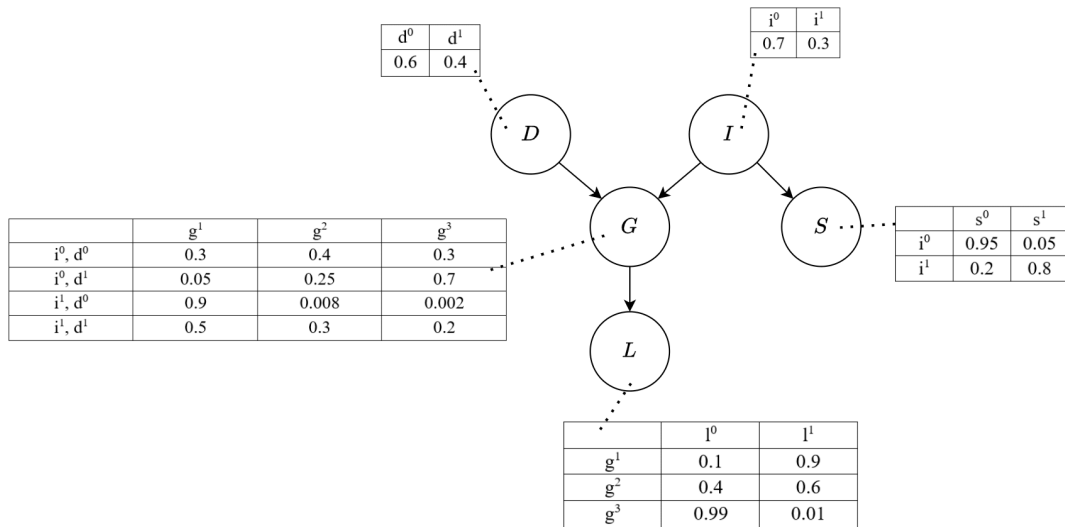


Figura 1.2: Xarxa Bayesiana sobre el model d'un estudiant amb les distribucions de probabilitat condicionades.

Els nodes del graf 1.2, que representen variables aleatòries, tenen assignada, o bé, una distribució de probabilitat (D i I segueixen una *Bernoulli* de paràmetres 0.6 i 0.3 respectivament) o bé, una distribució de probabilitat condicionada. Per exemple, la nota del curs G d'una persona participant dependrà la intel·ligència I i la dificultat D del curs. La probabilitat que tingui una bona nota $g_1 \in Val(G)$ si és considerada intel·ligent $i^1 \in Val(I)$ i el curs difícil és $d^1 \in Val(D)$ s'expressarà $P(G = g^1 \mid I = i^1, D = d^1)$, i en aquest cas valdrà 0.5. La probabilitat que tingui

una bona nota $g^1 \in Val(G)$ si és considerada intel·ligent $i^1 \in Val(I)$ i el curs és fàcil $d^0 \in Val(D)$, $P(G = g^1 \mid I = i^1, D = d^0) = 0.9$. La probabilitat que la qualitat de la carta L sigui bona $l^1 \in Val(L)$ si té una nota mitja g_2 és $P(L = l^1 \mid G = g_2) = 0.6$.

Del graf de la figura 1.2 per sí sol podem deduir independències del tipus: si saps que tant intel·ligent és una persona, la qualificació a l'examen SAT S és independent de la del curs G , ja que S només depèn de quant intel·ligent sigui la persona participant. Aquesta independència s'escriurà $(S \perp G \mid I)$. La distribució de probabilitat que segueixen les variables plantejades veu reflectida les independències del graf. És a dir, si saps el valor $i^k \in Val(I)$, llavors la distribució de probabilitat que el valor de l'examen $s_m \in (S)$ és independent de la nota $g_n \in Val(G)$ que pugui haver qualificat al curs:

$$P(S = s^m, G = g^n \mid I = i^k) = P(S = s^m \mid I = i^k) \cdot P(G = g^n \mid I = i^k)$$

Un avantatge dels models gràfics probabilístics és que proporcionen un marc de treball per facilitar la construcció efectiva de models que representen la realitat aprenent de les dades de la nostra experiència passada, i oferint una aproximació de les prediccions.

Més endavant, al capítol 4, entrarem més en detall veient la teoria que comprèn els PGMs.

1.3 Xatbots en comunitat virtuals

L'estudi bibliomètric sobre els agents conversacionals o xatbots [15] mostra la seva tendència creixent com a tema de recerca donada la seva àmplia aplicació. Tanmateix, els autors destaquen la fragmentació de casos d'ús reals, és a dir, cal analitzar on hi ha les necessitats de la comunitat virtual per enfocar una especialització de l'assistent virtual quan es vol garantir la seva prosperitat.

S'han realitzat diversos experiments que incentiven la implicació dels persones usuàries [20] o creen afecte [21]. És mitjançant mètodes de comportament adaptatiu quan un xatbot pot arribar a millorar notablement les experiències dels persones participants, involucrant-se en la col·laboració i la coproducció entre membres, considerant-lo un company més de la comunitat [23].

A l'hora d'implementar un xatbot s'han de tenir en compte les millores en les tasques d'acompanyament orientat a activitats específiques i d'aprofitament de l'ús d'un històric de converses. També cal fer millores de prediccions si s'utilitzen notificacions o actualitzacions contextuais. Existeixen reptes tècnics per tal que es els agents conversacionals es comportin com s'espera: entendre les percepcions, les expectatives i els contextos d'ús de les persones usuàries probablement impulsarà que sigui una eina més eficaç [22].

Per que un agent conversacional sigui capaç de comprendre l'entorn, el perfil de persones participants que conformen la comunitat, i altres aspectes relacionats amb el tipus de comunitat, en aquest projecte es proposa un model d'aprenentatge autònom basat en el procés d'inferència d'un PGM, fent possibles respondre'ns

preguntes com “*Quina proposta podria encaixar millor donat un usuari? Quin perfil de proposta genera més repercussió? Quin perfil d’usuari té més capacitat d’influència? Quina opinió podria generar un usuari donada una intervenció?*”. Respondre aquestes qüestions donarà peu a crear noves formes d’interacció dins la comunitat i, probablement, més participació.

1.4 Objectius

L’objectiu general del treball consisteix en formalitzar el concepte de comunitat virtual i especificar un xatbot dins la comunitat per tal de millorar l’experiència de la persona usuària, desenvolupant un model aplicable per generar més interacció. Aquest objectiu es pot especificar en una sèrie d’objectius parcials més concrets:

1. Formalitzar el concepte de comunitat virtual, considerant propostes similars en la literatura científica. Fer una interpretació de la utilitat d’un xatbot en una comunitat virtual, passant per formalitzar la seva figura.
2. Exemplificar aquesta formalització en el context de dues comunitats virtuals (centrades en educació i en participació ciutadana), i proposar diversitat d’interaccions amb el xatbot.
3. Desenvolupar un mètode que permeti estimar la incertesa associada a les preferències dels persones participants en un cas pràctic: la comunitat virtual de `decidim.barcelona`. Ha de ser un model aplicable a un xatbot, que pugui fer deduccions per tal d’atraure l’atenció dels persones usuàries i fomentar la interacció entre elles.

1.5 Estructura de la memòria

El present document s’estructura de la següent manera. Al capítol 2, es realitza un repàs en profunditat de l’estat de l’art, veient quins treballs formalitzen el concepte d’una comunitat virtual, o factors d’aquesta. A continuació, a l’apartat 3, es presenta un model aplicable de la comunitat virtual `decidim.barcelona`, per tal que un xatbot pugui generar més interacció aprofitant les dades observades prèviament, és a dir, dades d’interaccions de persones usuàries. El document acaba amb l’apartat 4, on s’exposen les conclusions extretes i una sèrie d’idees per estendre el treball en el futur.

Capítol 2

Estat de l'art

Són diversos els estudis que han portat les comunitats virtuals a un marc de treball més formal, enfocant-se cadascun des de diverses perspectives segons els interessos de cada projecte:

Per exemple, [6] utilitza l'especificació Z (un llenguatge utilitzat a enginyeria del software per l'especificació formal d'un sistema de còmput) per construir gradualment una sèrie de propietats desitjables en un marc que es podria utilitzar per especificar la jerarquia d'una comunitat: en primer lloc, veure les relacions que s'estableixen entre els membres de la comunitat i, en segon lloc, observar la pertinença dels membres a les diferents subcomunitats que es generen dins la comunitat. A partir d'aquí, es proposen una sèrie d'aplicacions i operacions per tal d'obtenir diferents resultats. Per exemple, inicia donant forma a un conjunt $Noms$ com un conjunt d'entitats, siguin persones o comunitats de persones. Parteix de la idea que la comunitat està formada per $Persones$ i $Comunitats$ de persones. La connexió de persones la representa la funció *connexió*: $Persones \leftrightarrow Persones$, i la pertinença a alguna $Comunitat$ ve representada per *pertinença*: $Persones \cup Comunitats \leftrightarrow Comunitats$. Aleshores, es poden construir tota una sèrie de funcions per deduir quines persones estan més allunyades, en el sentit que no tenen connexions amb altres persones participants o no pertanyen a cap comunitat. D'altra banda, si s'observa només les persones que tenen connexions es poden obtenir conjunts de comunitats que tinguin moltes connexions. L'objectiu d'aquest marc és utilitzar-lo com a base per a eines de visualització de comunitats.

En el cas de [11] es consideren les comunitats com construccions socials úniques que requereixen un procés subtil d'organització per ser sostenibles. És per aquest motiu que l'article es focalitza més en la governança de la comunitat i com aquesta evoluciona. Per tal que la comunitat prosperi es considera que els membres han d'adquirir un rol. Seran uns participants seleccionats els que tinguin “*converses per l'especificació*”, és a dir, converses sota el seu rol on podran dur a terme certes accions amb l'objectiu de prendre decisions. Per conceptualitzar-ho, utilitza normes de composició formal per seleccionar els membres rellevants. Parteix de la definició d'alguns conjunts, com SP , que contindrà tots els processos d'especificació, és a dir, creacions, modificacions o terminacions de definicions de coneixement, o DE el conjunt d'efectes deòntics, que poden ser *permès*, *prohibit*, *requerit*, segons el grau de

responsabilitats. També CP serà el conjunt de processos de control (les iniciacions, execucions i terminacions de processos) per especificar les converses. S'utilitza la norma de composició: $d_{cn} = \langle id, de, a, cp, sp \rangle$, on id serà un identificador i $a \in A$ correspondrà al rol que pot prendre, $de \in DE$, $cp \in CP$ i $sp \in SP$. Serà mitjançant aquesta norma de composició i diverses funcions, com la que assigna l'efecte deòntic a cada usuari segons les normes, el control de procés i el procés d'especificació, la que permeti conceptualitzar les converses per l'especificació, amb l'objectiu que es puguin adaptar a les necessitats de suport d'una comunitat.

També es considera rellevant l'article [7] que desenvolupa una base per a un model d'intercanvi d'informació en comunitats virtuals i proposa una tipologia de les varietats de comportament dels participants que es troben a les comunitats virtuals. En un inici, la tipologia separa el comportament no interactiu, com pot ser el participant passiu que només es dedica a llegir els texts creats per altres participants, de l'interactiu, i dins d'aquest es torna a dividir entre l'hostil, que involucra insults, troleig i publicitat de qualsevol tipus, entre d'altres, i el comportament interactiu positiu. Aquest últim incorpora tant activitats específicament relacionades amb la cerca i compartició d'informació, com activitats més socials. Aquesta tipologia proporcionarà un mecanisme per avaluar les característiques de les comunitats virtuals pel que fa al seu suport a l'intercanvi d'informació, tenint potencial de millorar la comprensió de les comunitats virtuals com a entorns d'informació.

La tesi [4] estudia els serveis d'assistència en entorns híbrid-virtuals estructurats. Per fer-ho presenta la formalització d'aquests entorns. En un principi diferencia dues capes: per una banda, una capa organitzativa on es representa la figura dels persones usuàries, els objectius i les interaccions que tenen. Per altra, la capa d'assistència a les persones usuàries, que es caracteritza per comptar amb assistents, cadascun dels quals proveirà un servei al sistema. La primera d'elles es formalitza com un tupla $Org = \langle O, SocStr, SocConv, DomP, Goals, AssQoSi \rangle$ on cada element caracteritza l'entorn. Per exemple, el conjunt O conceptualitza l'ontologia de l'espai, comprenent des dels conceptes més simples fins els més complexos, i les relacions entre aquests. L'estructura social $SocStr$ formalitza les relacions i els rols que prenen els membres. És a les convencions socials $SocConv$ on entra el concepte d'activitat i les accions que poden prendre els participants segons el rol que desenvolupen en les pròpies activitats, així com també apareixen les normes i els protocols per la compleció d'activitats. Aquests elements en concret seran els que es desenvolupen similarmet dins el marc d'una comunitat virtual.

Capítol 3

Formalització d'una comunitat virtual

Amb l'objectiu de definir un escenari per poder treballar sobre les comunitats virtuals, es proposa en primer lloc una formalització d'aquestes, conceptualitzant aquelles variables que siguin necessàries per tal que sigui un model aplicable a qualsevol tipus de comunitat. Posteriorment s'introduirà el cas pràctic de dos exemples: una comunitat virtual educativa i una centrada en processos participatius. A més, veurem com ens pot beneficiar la integració d'un xatbot en la comunitat virtual i, a partir d'aquí, veurem com el podríem incorporar segons la nostra formalització.

3.1 Proposta

En aquesta secció farem una proposta de modelització d'una comunitat virtual, observant cada variable que defineix una comunitat virtual, considerant els diferents espais d'interacció i les activitats que sorgeixen dins la virtualitat. Iniciarem donant una sèrie de definicions amb l'objectiu de descriure el model per les comunitats virtuals, podent-se aplicar a qualsevol tipus de comunitats:

Definició 3.1.1. Una *Comunitat Virtual* està formada per una ontologia (O), una estructura social (ES), un espai virtual (EV) i unes convencions socials (CS):

$$CV = \langle O, ES, EV, CS \rangle$$

Abans de veure en què consisteix cada terme d'aquesta definició, i per tal de definir posteriorment algunes funcions, recordem la definició del conjunt potència $\mathcal{P}(A)$ d'un conjunt A qualsevol:

Definició 3.1.2. El conjunt potència de A és el conjunt $\mathcal{P}(A) = \{x \mid x \subseteq A\}$, és a dir, és el conjunt de subconjunts de A .

Un cop recordada aquesta definició, veiem què caracteritzen cadascun dels elements de la definició de **comunitat virtual**.

La primera component que defineix la comunitat virtual és l'**ontologia**. En el camp de la tecnologia, l'ontologia [2] és un mecanisme de representació en forma d'esquema conceptual que formula una especificació formal dels conceptes continguts en un domini o àrea de coneixement abastint des dels més simples fins els més complexos, donant significat a cadascun d'ells i a les relacions entre aquests:

Definició 3.1.3. Una *Ontologia* és la 5-tupla

$$O = \langle Con, Def, Dic, R_C, ConRel \rangle$$

on:

- El conjunt *Con* contindrà tots els conceptes que tinguin a veure amb la temàtica de la comunitat.
- La definició de cadascun dels conceptes estarà continguda al conjunt *Def*.
- Donat un concepte, hi haurà la correspondència amb la definició mitjançant la funció

$$Dic : Con \longrightarrow Def$$

- El conjunt *R_C* contindrà les relacions que s'estableixen entre els diferents conceptes.
- Per tal d'obtenir les relacions que s'estableixen entre dos conceptes s'utilitza la funció

$$ConRel : Conceptes \times Conceptes \longrightarrow R_C \cup \{\emptyset\}$$

L'**estructura social** constitueix la segona component de l'especificació d'una comunitat virtual. Per una banda, formalitza el conjunt de relacions i interaccions socials que es donen entre els diferents membres d'una determinada comunitat. A [6] s'obté un marc de treball sobre les comunitats virtuals per poder entendre les subcomunitats que es formen mitjançant la participació i intervenció en diferents activitats. Totes les subcomunitats engloben subcomunitats de persones que participen a diferents espais d'interacció. Podem, per tant, saber les connexions que té una subcomunitat *S* a partir de les connexions que tenen les seves subcomunitats $S_1, S_2 \subset S$, i així reiteradament. Aquest concepte es coneix subcomunitats per capes, i entendre-les permetrà extreure conclusions de les relacions que es generen entre individus.

A [4] es consideren els diferents rols que els individus poden prendre a la comunitat, així com les propietats de cada membre i subcomunitat. La proposta d'assignació de rols i accions també es reforça a [11] per tal que la comunitat evolucioni i s'adapti. Els rols [3] corresponen al conjunt de comportaments o pautes de conducta que s'esperen dels individus en una situació social determinada. Es tracta d'un comportament que pot ser esperat, lliure o estar en constant canvi i pot tenir un determinat estatus social o posició social individual. Entendrem:

Definició 3.1.4. L'*Estructura Social* està definida pels membres (Mem), les relacions (Rel_M) que s'estableixen, les subcomunitats ($SubCom$) que es formen, les propietats ($Prop_{M-SC}$) d'aquestes i la seva caracterització ($Carac_{M-SC}$), i els rols (Rol) que existeixen:

$$ES = \langle Mem, Rel_M, SubCom, Prop_{M-SC}, Carac_{M-SC}, Rol \rangle$$

on:

- El conjunt Mem és el conjunt de membres de la comunitat.
- La relació que té un membre d'una comunitat amb altres membres de la mateixa vindrà representada per la funció

$$Rel_M : Mem \longrightarrow \mathcal{P}(Mem)$$

- En diverses ocasions s'han de considerar agrupacions de membres amb un propòsit, per exemple, quan s'agrupen per afrontar un projecte. Per això, considerarem les subcomunitats $SubCom$ com subconjunts de Mem , és a dir:

$$S \subseteq Mem, \forall S \in SubCom$$

- Normalment, tant els membres com les subcomunitats tenen una sèrie d'atributs que descriuen el seu perfil. Aquestes propietats estaran contingudes a un conjunt $Prop_{M-SC}$, el qual pot variar segons el tipus de comunitat i els seus interessos. Poden ser des del nom o gènere de cada membre fins el propòsit d'una subcomunitat.
- Per tal de caracteritzar cada membre o subcomunitat amb les seves propietats del seu perfil, s'utilitzarà la correspondència:

$$Carac_{M-SC} : Mem \cup SubCom \longrightarrow \mathcal{P}(Prop_{M-SC})$$

- El conjunt Rol és el conjunt de totes les possibles posicions socials que desenvolupen tasques dins la comunitat. Per exemple, en una comunitat educativa existeix el rol de l'estudiant i el del professor. Veurem més endavant com es connecten els comportaments o accions permeses amb el rol que es desenvolupa.

La tercera component que defineix una comunitat virtual és l'**espai virtual**. És el conjunt que modelitza les activitats en les quals es pot participar, cadascuna caracteritzada per les seves propietats. Cada comunitat virtual tindrà la seva forma específica d'identificar les activitats i assignar-les propietats. A [7] s'estudia el comportament interactiu de les persones usuàries i, mitjançant aquest, realitza una classificació de les activitats. També és a l'espai virtual on trobarem els rols que es

poden desenvolupar segons l'activitat i el membre, i les possibles accions a dur a terme en una activitat donat un rol.

L'**espai virtual** l'entendem com el conjunt que modelitza les activitats informatives, discursives i participatives, així com els rols s'hi desenvolupen i les accions que estan a l'abast de cada usuari a les activitats que participa. Formalment tindrem:

Definició 3.1.5. L'*Espai Virtual* està definit per les activitats (Act) que s'hi desenvolupen, les accions que podran prendre (Acc), les propietats ($Prop_A$) de les activitats i la seva caracterització ($Carac_A$), l'assignació (Rol_{AM}) de rols donat un activitat i un membre, i l'assignació (Acc_{AR}) d'accions possibles donada una activitat i un rol:

$$EV = \langle Act, Acc, Prop_A, Carac_A, Rol_{AM}, Acc_{AR} \rangle$$

on:

- El conjunt Act és el conjunt d'activitats, espais d'interacció entre els diferents membres de la comunitat. Són en elles on els usuaris tindran a l'abast algunes accions a dur a terme per involucrar-se en la pròpia comunitat. Aquest conjunt varia segons la classe de comunitat: poden anar des d'una avaluació o una classe a una comunitat educativa, fins una votació o un debat a una comunitat basada en processos participatius.
- El conjunt Acc conté totes les possibles accions que poden realitzar els membres quan prenen un rol en les activitats.
- El conjunt $Prop_A$ contindrà tots les propietats de les activitats, com ara el tipus i el contingut que tracten. També serà un conjunt que variarà segons els interessos de cada comunitat virtual.
- Per tal de caracteritzar cada activitat amb les seves propietats, les subcomunitats que hi participen i els conceptes que involucra, s'utilitzarà la correspondència:

$$Carac_A : Act \rightarrow \mathcal{P}(Prop_A) \times \mathcal{P}(SubCom) \times \mathcal{P}(Con)$$

A més, les activitats també es caracteritzen per les accions que es donen i els rols que hi participen:

- Per tal d'obtenir el rol que pot prendre un membre donada una activitat emprarem la funció:

$$Rol_{AM} : Act \times Mem \longrightarrow Rols \cup \{\emptyset\}$$

- Si donat un rol volem saber les accions possibles en una activitat prendrem la funció:

$$Acc_{AR} : Act \times Rols \longrightarrow \mathcal{P}(Acc)$$

Part del comportament interactiu dins d'una comunitat virtual ve limitat per una sèrie de **convencions socials**, la quarta component de la tupla que defineix la comunitat virtual. Determinen el conjunt d'estàndards, regles, normes o també criteris que són d'acceptació general per a un determinat grup social [1]. Seguint els passos de [4], les persones participants prenen un rol cada cop que participen en les activitats, on completen les tasques seguint uns protocols, o es mouen a altres activitats relacionades seguint les restriccions definides. Veiem com es pot formalitzar:

Definició 3.1.6. Les *Convencions Socials* és la terna

$$CS = \langle Norm, Prot, Prot_{AN}, ActRel \rangle$$

on:

- El conjunt *Norm* és el conjunt de normes i principis que s'imposen o s'adopten per dirigir la conducta o la correcta realització d'una acció o el correcte funcionament de les activitats. Són inherents i s'apliquen sobre cada membre de la comunitat.
- El conjunt *Prot* contindrà tots els protocols de realització de tasques, és a dir, totes les seqüències per dur a terme la compleció de totes les activitats
- Amb la finalitat de saber quins protocols s'han de seguir per la compleció d'activitats es farà ús de la funció

$$Prot_{AN} : Act \longrightarrow Prot \cup \{\emptyset\}$$

- Sovint s'utilitzen diverses combinacions d'activitats per tal de complir amb un objectiu. Per exemple, si tractem una comunitat educativa podria ser el cas de diferents avaluacions d'un mateix curs, o en el cas d'una comunitat de participació ciutadana, les diferents propostes pertanyents a un mateix procés participatiu. Per tal de saber les relacions entre activitats, considerarem la funció

$$ActRel : Act \longrightarrow \mathcal{P}(Act)$$

Segons el tipus d'activitats involucrades, l'ordre del conjunt sortida serà rellevant o no.

Fins aquí s'ha vist que comprèn la definició d'una comunitat virtual. A continuació, es proposa l'extensió d'una comunitat virtual amb la formalització de la figura del xatbot.

Funcionalitats i propietats d'un xatbot

Els xatbots utilitzen llenguatge natural per comunicar-se amb els usuaris, el que fa que siguin sistemes socials que exhibeixen certes característiques humanes (gènere, personalitat, estat emocional, aparença). Per exemple a [12] es veu que un xatbot es percep de manera més positiva quan l'usuari forma la impressió d'un agent del mateix gènere.

A causa que poden socialitzar amb humans, com els humans els perceben és un tema important. Hi ha estudis que diuen que els humans interactuen amb els agents conversacionals com ho fan amb les seves mascotes [19], és a dir, com a companys. També es va realitzar un estudi [8] sobre com els usuaris veien els futurs robots acompanyants. Els resultats van estar: el 79% dels usuaris els perceben com assistents, el 71% com a màquines, el 46% com a servents i un percentatge menor els consideraven amics. La relació entre els xatbots i els humans pot variar per segons els usuaris. Per tant, com un xatbot es dirigeix a un usuari ha de ser dissenyat en concordança amb la visió que els usuaris tenen dels xatbots [5].

A l'hora de dissenyar un agent conversacional es tenen en compte diversos factors que poden influenciar en la percepció de l'usuari. Aquests són:

- Característiques antropomòrfiques com aparença, així com nom o, fins i tot, gènere (ser binari o no binari, sense gènere, etc.). També la personalitat que poden tenir (neutre, amigable, seriosa, etc.).
- L'estil de llenguatge que utilitza, formal o informal.
- Si és adaptatiu en quant que canviï al llarg del temps adaptant l'estil del llenguatge segons la persona usuària, el contingut del diàleg, o, fins i tot, el rol que pren i les responsabilitats que se l'assignen.
- Les plataformes de contacte que s'ofereixen: si els usuaris es comuniquen amb l'agent a través d'una plataforma (monoplataforma) o des de diverses (multiplataforma).
- La modalitat, és a dir, com es realitza la comunicació amb els usuaris (veu, text, reaccions, etc.).
- La gestió de la privacitat dels membres de la comunitat.

Per tal de generar més interacció dins una comunitat virtual es proposa la incorporació d'un **xatbot**, que mantingui diàlegs amb persones usuàries. Aquests diàlegs poden estar basats en l'acompanyament d'activitats o, simplement, poden donar-se amb l'objectiu de socialitzar, dotant al xatbot uns tòpics dels que sàpiga parlar. Com és habitual, cadascun dels xatbots tindrà una caracterització: les propietats particulars, els conceptes que abasta, les activitats que acompanya i els membres amb els quals es relaciona. A més, els xatbots poden tenir memòria per poder recordar certes experiències que li faciliten mantenir el context de la conversa entre xatbots i persones involucrades, amb el propòsit de garantir una bona experiència

d'usuari. Alhora, un xatbot pot agafar un rol donada una activitat i podrà prendre accions segons el rol que se l'assigni.

Per tal d'encaixar el xatbot dins del model de comunitat virtual plantejat prèviament considerarem la definició:

Definició 3.1.7. Un xatbot X està definit per les propietats ($Prop_X$) que obté a les activitats, els diàlegs (Dlg) que manté amb diferents membres, la seva caracterització ($Carac_D$) dels diàlegs, mitjançant la qual establim el conjunt de membres, activitats i conceptes que abasta, la memòria (M_X) dels diàlegs amb cada membre, els rols (Rol_X) que pren dins les activitats, i les accions (Acc_X) que poden dur a terme:

$$X = \langle Prop_X, Dlg, M_X, Carac_D, Rol_X, Acc_X \rangle$$

on:

- El conjunt $Prop_X$ serà el conjunt de les diferents propietats assignades al xatbot.
- El conjunt Dlg serà el conjunt de diàlegs que estableix el xatbot amb la persona participant.
- Per tal de garantir una bona experiència d'usuari mantenint el context dels diàlegs que es donen entre xatbots i persones usuàries, el xatbot memoritzarà determinades parts dels diàlegs. Aquest conjunt s'anomenarà M_X , caracteritzarà el diàleg i ajudarà en les preses de decisions del xatbot.
- Els diàlegs vindran caracteritzats pels conceptes amb els que el xatbot sàpiga mantenir un diàleg, si té a veure amb l'acompanyament de tasques les activitats que comprèn, les parts del diàleg memoritzades, i, finalment, el subconjunt de membres amb els que dialoga:

$$Carac_D : Dlg \longrightarrow \mathcal{P}(Con) \times \mathcal{P}(Act) \times \mathcal{P}(M_X) \times \mathcal{P}(Mem)$$

- Donada una activitat, el rol i les propietats que defineixin la figura del xatbot que pugui desenvolupar vindran modelades per la funció:

$$Rol_X : Act \longrightarrow \mathcal{P}(Rols) \times \mathcal{P}(Prop_X)$$

- Segons una activitat, el rol que el xatbot pugui prendre i la memòria de la que disposi, les accions permeses per prendre s'especifiquen amb la funció:

$$Acc_X : Act \times Rols \times M_X \longrightarrow \mathcal{P}(Acc)$$

La seqüència d'accions que podria dur a terme un xatbot seria $Acc_1 = \{ \text{“entendre la locució (o missatge) de la persona interlocutora”, “trobar una resposta al missatge”, “consultar en sistemes externs per poder respondre a la persona participant (consultar una base de dades, o un servei web, per exemple per dir a l'usuari el temps que farà demà)”} \subset Acc$.

La següent definició conceptualitza l'extensió d'una comunitat virtual amb un xatbot:

Definició 3.1.8. Una *Comunitat Virtual enriquida amb un Xatbot* està formada per una comunitat virtual (CV) i un xatbot (X), els quals comparteixen persones participants amb les que el xatbot dialoga, activitats que es desenvolupin a la comunitat virtual, conceptes amb els que el xatbot pugui mantenir un diàleg, etc.

$$CV_X = \langle CV, X \rangle$$

Amb aquests conceptes definits ja podem començar a construir exemples on es vegi una aplicació d'aquesta formalització.

3.2 Exemples

En aquesta secció veurem un parell d'exemples reals de comunitats virtuals, les quals agafarem i formalitzarem mitjançant la proposta de l'apartat anterior.

3.2.1 Comunitats Virtuals en l'educació

Els cursos en línia oberts massius (MOOCs, per les sigles en anglès) estan proliferant ràpidament. Segons Class Central [24], l'any 2016 es van registrar més de 58.000.000 d'estudiants de tot el món per a més de 6.800 cursos oferts per més de 700 institucions. A més, aquestes xifres continuen creixent ràpidament. Avui en dia, aquests cursos cobreixen gairebé totes les disciplines i nivells educatius.

En aquest apartat considerarem el curs online “*Alimentación y dietética para una vuelta al mundo a vela*” ofert per la Universitat de Barcelona a la plataforma de *Coursera*. A més, també s'afegiran situacions d'interacció proporcionades per la plataforma que no ofereix el curs, però que resulten igual de vàlides. L'**Ontologia**

de la comunitat virtual comprendrà tots aquells conceptes que girin al voltant de la fisiologia de l'esport, la nutrició i la dietètica, l'esport transoceànic, la hidratació i la ciència i tecnologia dels aliments.

Per exemple, el curs comença amb les definicions “el procés d'entrenament és l'administració de càrregues de treball que produiran una sobrecàrrega biològica de l'organisme i els canvis metabòlics són la resposta fisiològica al procés d'entrenament que determina la progressió del rendiment de l'esportista”. També es considera tots aquells conceptes relacionats amb el món de l'educació, com avaluació o classe. Dins la formalització tindrem:

$$\begin{aligned} con_1 &= \text{“entrenament”}, con_2 = \text{“canvi metabòlic”}, \\ con_3 &= \text{“classe”}, con_4 = \text{“avaluació”}, \\ con_1, con_2, con_3, con_4 &\in Con \end{aligned}$$

$$def_1 = \text{“administració de càrregues de treball que produiran una”}$$

$$\begin{aligned}
& \text{sobrecàrrega biològica del organisme"}, \\
def_2 &= \text{"resposta fisiològica al procés d'entrenament que determina} \\
& \quad \text{la progressió del rendiment de l'esportista"}, \\
def_3 &= \text{"conjunt d'alumnes que segueixen una mateixa assignatura i} \\
& \quad \text{assisteixen a les lliçons corresponents"}, \\
def_4 &= \text{"etapa del procés educatiu que té per finalitat comprovar de quina} \\
& \quad \text{manera han estat aconseguits els objectius pedagògics proposats"}, \\
& \quad def_1, def_2, def_3, def_4 \in Def \\
Dic(con_1) &= def_1, Dic(con_2) = def_2, Dic(con_3) = def_3, Dic(con_4) = def_4
\end{aligned}$$

Fixant-nos també la relació que tenen entre els conceptes proposats podem dir que el canvi metabòlic és la resposta fisiològica al procés d'entrenament i, per exemple, que una classe segueix una avaluació:

$$\begin{aligned}
relc_1 &= \text{"resposta fisiològica a"} \in R_C, \\
relc_2 &= \text{"segueix"} \in R_C, \\
ConRel(con_2, con_1) &= relc_1, ConRel(con_3, con_4) = relc_2
\end{aligned}$$

Veiem en què consisteix l'**estructura social** de la comunitat virtual de la plataforma *Coursera*, que és on pertany el curs ofert. Podem considerar tot el conjunt *Mem* de membres de tota la plataforma i $S_1, \dots, S_n \in SubCom$, com els conjunts de membres que coincideixen fent activitats o cursos. Podríem considerar les relacions dels membres amb altres persones participants:

$$Rel_M(m) \subseteq \bigcup_{\substack{1 \leq i \leq n, \\ S_i \ni m}} S_i, \quad m \in Mem$$

En quant a les propietats, els usuaris que estan a la plataforma poden completar el seu perfil amb una sèrie d'atributs com el nom, la ubicació, els interessos personals, etc. Les subcomunitats poden tenir un objectiu o un nombre màxim de persones, per exemple. Encaixant-ho amb la formalització:

$$\begin{aligned}
& nom, ubicació, interessos, objectiu, màxim \in Prop_{M-SC}, \\
\{nom, ubicació, interessos\} &\subseteq Carac_{M-SC}(m), m \in Mem, \\
\{nom, objectiu, màxim\} &\subseteq Carac_{M-SC}(s), s \in SubCom
\end{aligned}$$

En un inici, a la plataforma educativa es poden trobar dos rols principals: l'instructor i l'estudiant. No obstant, compta també amb assistents, que donen suport a l'instructor a l'hora de respondre dubtes, i els avaluadors, que permeten fer la tasca de correcció més ràpida. Per tant, queda:

$$\begin{aligned}
rol_1 &= \text{"instructor"}, rol_2 = \text{"estudiant"}, rol_3 = \text{"assistent"}, \\
rol_4 &= \text{"avaluador"}, rol_1, rol_2, rol_3, rol_4 \in Rols
\end{aligned}$$

S'observa també l'**espai virtual** que podria definir la diversitat d'activitats que s'ofereixen. Per una banda tindriem les activitats informatives que correspondrien a les que tractessin contingut educatiu (classes, apunts, vídeos, etc.), les activitats discursives que serien aquelles en què dos o més usuaris entaulen una conversa (debat, fòrums, etc.) i les activitats participatives, aquelles que requereixen la participació dels usuaris (votació, avaluació, etc.). Les activitats podrien ser tant síncrones (videoconferències en línia o classes en directe) com asíncrones (apunts descarregables o classes gravades), i serà en elles també on es podran donar accions de l'usuari (respondre un missatge en un fòrum, descarregar un arxiu, reproduir el contingut, etc.). Diversos exemples respecte la formalització serien:

$$act_1 = \text{"fòrum"}, act_2 = \text{"apunts"}, act_3 = \text{"classe virtual en línia"}, act_4 = \text{"projecte"}, \\ act_1, act_2, act_3, act_4 \in Act$$

$$acc_1 = \text{"publicar missatge"}, acc_2 = \text{"respondre un missatge"}, acc_3 = \text{"enviar comunicat"}, acc_4 = \text{"visualitzar"}, acc_5 = \text{"descarregar contingut"}, \\ acc_6 = \text{"unir-se"}, acc_7 = \text{"preguntar"}, acc_8 = \text{"abandonar"}, acc_9 = \text{"penjar document"}, acc_{10} = \text{"qualificar"}, acc_{11} = \text{"editar contingut"}, acc_{12} = \text{"compartir contingut en directe"}, acc_1, \dots, acc_{12} \in Acc$$

$$nom, tipus, sincronia, caràcter, qualificable \in Prop_A$$

Si considerem $S_i \subset Subcom$ la subcomunitat que realitza l'activitat act_i i $C_i \subset Con$ el subconjunt de conceptes que involucra, una construcció de la funció $Carac_A$ seria:

$$Carac_A(act_1 = \text{"fòrum"}) = (\{nom, tipus = \text{"discursiva"}, sincronia = \text{"Cert"}\}, S_1, C_1) \\ Carac_A(act_2 = \text{"apunts"}) = (\{nom, tipus = \text{"informativa"}, sincronia = \text{"Fals"}, caràcter = \text{"individual"}\}, S_2, C_2) \\ Carac_A(act_3 = \text{"classe en línia"}) = (\{nom, tipus = \text{"informativa participativa"}, sincronia = \text{"Cert"}\}, S_3, C_3) \\ Carac_A(act_4 = \text{"projecte"}) = (\{nom, tipus = \text{"participativa"}, caràcter = \text{"col·laboratiu"}, qualificable = \text{"Cert"}\}, S_4, C_4)$$

Es pot ser tant específic com sigui necessari ampliant el conjunt de propietats que identifiquen les activitats. La propietat nom serà particular per cada activitat d'una comunitat virtual, en el cas d'un projecte de desenvolupar un $nom = \text{"model d'entrenament i nutrició eficaç per navegants de vela"}$, aleshores podria anomenar-se nom .

Per assignar els rols que podran prendre els membres és necessària l'obtenció de la funció Rol_{AM} . Donat $mem \in Mem$ veiem quins rols es poden prendre i quines accions poden desenvolupar aquests rols:

- Per $act_1 = \text{"fòrum"}$ es podran desenvolupar dos rols: el d'estudiant, que pot preguntar o respondre missatges, i el d'assistent, que a més de publicar o respondre missatges, podrà enviar un comunicat.

$$\begin{aligned} \text{Rol}_{AM}(\text{act}_1, \text{mem}) &\in \{\text{rol}_2, \text{rol}_3\} \\ \text{Acc}_{AR}(\text{act}_1, \text{rol}_2) &= \{\text{acc}_1, \text{acc}_2\} \\ \text{Acc}_{AR}(\text{act}_1, \text{rol}_3) &= \{\text{acc}_1, \text{acc}_2, \text{acc}_3\} \end{aligned}$$

- Per $\text{act}_2 = \text{“apunts”}$ es podran desenvolupar el rol d’estudiant, el d’instructor i el d’assistent. L’estudiant podrà visualitzar el contingut o descarregar-lo, mentre que l’instructor i l’assistent, a més d’aquestes accions, podran també editar contingut.

$$\begin{aligned} \text{Rol}_{AM}(\text{act}_2, \text{mem}) &\in \{\text{rol}_1, \text{rol}_2, \text{rol}_3\} \\ \text{Acc}_{AR}(\text{act}_2, \text{rol}_1) = \text{Acc}(\text{act}_2, \text{rol}_3) &= \{\text{acc}_4, \text{acc}_5, \text{acc}_{11}\} \\ \text{Acc}_{AR}(\text{act}_2, \text{rol}_2) &= \{\text{acc}_4, \text{acc}_5\} \end{aligned}$$

- Per $\text{act}_3 = \text{“classe en línia”}$ es podran desenvolupar el rol d’estudiant i el d’instructor. Tant l’estudiant com l’instructor podran unir-se o abandonar la sala, l’estudiant podrà preguntar i l’instructor podrà compartir contingut en directe.

$$\begin{aligned} \text{Rol}_{AM}(\text{act}_3, \text{mem}) &\in \{\text{rol}_1, \text{rol}_2\} \\ \text{Acc}_{AR}(\text{act}_3, \text{rol}_1) &= \{\text{acc}_6, \text{acc}_8, \text{acc}_{12}\} \\ \text{Acc}_{AR}(\text{act}_3, \text{rol}_2) &= \{\text{acc}_6, \text{acc}_7, \text{acc}_8\} \end{aligned}$$

- Per $\text{act}_4 = \text{“projecte”}$ es podrà desenvolupar el rol d’estudiant i el d’avaluador. Per una banda, l’estudiant podrà penjar un document o editar-ne el contingut. D’altra banda, l’avaluador podrà descarregar el contingut i qualificar-lo.

$$\begin{aligned} \text{Rol}_{AM}(\text{act}_4, \text{mem}) &\in \{\text{rol}_2, \text{rol}_4\} \\ \text{Acc}_{AR}(\text{act}_4, \text{rol}_2) &= \{\text{acc}_9, \text{acc}_{11}\} \\ \text{Acc}_{AR}(\text{act}_4, \text{rol}_4) &= \{\text{acc}_5, \text{acc}_{10}\} \end{aligned}$$

Les **convencions socials** seran similars als cursos impartits presencialment a les institucions educatives. El que els distingirà de les presencials serà les normes d’ètica i privacitat, ja que estem tractant del món en línia. Dos exemples de normes podrien ser $\text{norm}_1 = \text{“No pugis fotografies o vídeos d’altres sense el seu consentiment”}$ i $\text{norm}_2 = \text{“El plagi als exàmens i treballs és una infracció la sanció de la qual és la pèrdua de la convocatòria de l’estudiant”}$, $\text{norm}_1, \text{norm}_2 \in \text{Norm}$.

Per tal de completar una activitat és necessari seguir un protocol. Per exemple, per dur a terme un projecte, primer t’has de registrar i després es podrà penjar. També es consideraran els protocols davant l’incompliment d’una norma. Un exemple d’ambdós casos seria

$$\begin{aligned} \text{prot} &= \text{“registrar usuari, penjar document, avaluar contingut”} \in \text{Prot}, \\ \text{Prot}_{AN}(\text{act}_4) &= \text{prot} \end{aligned}$$

Si per exemple suposem que un mètode d'avaluació com un projecte act_4 que té darrere totes les referències que prèviament s'han penjat a la plataforma en forma d'apunts act_2 i classes virtuals en línia act_3 . A més, s'habilita un fòrum act_1 per resoldre dubtes del projecte. Podrem considerar llavors

$$ActRel(act_4) = \{act_1, act_2, act_3\}$$

Donada la comunitat virtual educativa, una possible incorporació d'un **xatbot** podria consistir en diferents interaccions a les activitats. Sense tenir en compte la memòria de diàlegs, es proposen diverses situacions per tal d'exemplificar la formalització prèvia del xatbot X :

Dins un $act_1 = \text{"fòrum"}$ d'un curs que pren la subcomunitat $S \in SubCom$ de persones estudiants, el xatbot actuaria com a dinamitzador enviant missatges que realcessin la importància del tema $C \subset Con$ tractat en el fil del fòrum, incentivant així la participació de les persones estudiants. Els diàlegs $dlg \in Dlg$ que es mantenguessin vindria caracteritzat per tots aquests factors, a més es considerarà $M \in M_X$ les parts memoritzades de dlg . El xatbot, exhibiria característiques antropomòrfiques com nom o gènere, amb el tipus de modalitat i la quantitat de plataformes que s'ofereixen:

$$\begin{aligned} nom &= \text{"Pep"}, gènere = \text{"home"}, modalitat = \text{"text"}, \\ &plataforma = \text{"monoplataforma"}, \\ &nom, gènere, modalitat, plataforma \in Prop_X, \\ &Carac_D(dlg) = (C, \{act_1\}, M, S) \\ (rol_1 &= \text{"dinamitzador"}, \{nom, gènere, modalitat, plataforma\}) \in Rol_X(act_1), \\ Acc_X(act_1, rol_1) &= Acc_1 \cup \{\text{"realçar la temàtica"}, \text{"enviar missatge"}, \text{"respondre missatge"}\} \subset Acc \end{aligned}$$

La incorporació d'un xatbot amb propietats més socials per tal de connectar persones usuàries i potenciar la interacció. Comptaria amb una personalitat i es consideren propietats com estil de llenguatge informal, que es mogui per diverses plataformes, etc. S'involucraria en diversitat d'activitats $act_1, \dots, act_n \in Act$, contactant amb tots els membres i mantenint diàlegs $dlg \in Dlg$ sobre tòpics $C \subset Con$ amb parts memoritzades $M \in M_X$:

$$\begin{aligned} nom &= \text{"Carles"}, plataforma = \text{"multiplataforma"}, personalitat = \\ &\text{"amigable"}, estil = \text{"informal"}, \\ &personalitat, estil \in Prop_X, \\ &Carac_D(dlg) = (C, \{act_1, \dots, act_n\}, M, Mem) \\ (rol_2 &= \text{"socialitzador"}, \{nom, plataforma, personalitat, estil\}) \in Rol_X(act_i) \forall 1 \leq \\ &i \leq n, \\ Acc_X(act_1, rol_2, M) &= Acc_1 \cup \{\text{"anàlisi perfil d'usuari"}, \text{"enviar missatge"}, \text{"crear fil"}, \text{"respondre missatge"}\} \subset Acc \end{aligned}$$

També podem considerar propietats més autoritàries i formals que actués en fòrums act_1, \dots, act_n de tota la comunitat. És el cas d'un estudi [7] on desenvolupen

un agent conversacional que intervé en la invisibilització d'aquells fils o missatges hostils que es puguin dur a terme, ja que un tipus d'interacció interpersonal són els atacs a altres membres de la comunitat, com fer publicitat o publicar contingut inapropiat. Els conceptes $C \subset Con$ que el caracteritzarien els diàlegs $dlg \in Dlg$ serien de caràcter normatiu amb les parts memoritzades $M \in M_X$. La incorporació del xatbot podria frenar ràpidament la proliferació d'aquest mal comportament:

$$\begin{aligned} &personalitat = \text{“autoritària”}, estil = \text{“formal”}, \\ &Carac_D(dlg) = (C, \{act_1, \dots, act_n\}, M, Mem) \\ &(rol_3 = \text{“intervencionista”}, \{personalitat, estil\}) \in Rol_X(act_i) \forall 1 \leq i \leq n, \\ &Acc_X(act_i, rol_3, M) = Acc_1 \cup \{\text{“analitzar missatge”}, \text{“amagar contingut”}, \text{“notificar usuari”}\} \subset Acc, \forall 1 \leq i \leq n \end{aligned}$$

Una altra manera de poder considerar un un agent és com informador o assistent en un fòrum (act_1). A [14] es veu els efectes d'utilitzar diversos agents dins una comunitat virtual educativa responent preguntes dels usuaris amb la finalitat d'alliberar càrrega de feina als treballadors. També podria comprendre el tancament de fils resumint l'activitat que s'ha donat. Els diàlegs $dlg \in Dlg$ amb les persones usuàries $m \in S \in SubCom$ vindran caracteritzats pels conceptes $C \subset Con$ que abasti la temàtica de l'activitat, amb les parts del diàleg dlg memoritzades $M \in M_X$:

$$\begin{aligned} &nom = \text{“Maria”}, gènere = \text{“dona”}, personalitat = \text{“neutre”} \\ &Carac_D(dlg) = (C, act_1, M, S) \\ &(rol_3 = \text{“assistent”}, \{nom, gènere, personalitat\}) \in Rol_X(act_1), \\ &Acc_X(act_1, rol_3, M) = \{\text{“anàlisi de fil”}, \text{“respondre pregunta”}, \text{“tancar fil”}\} \subset Acc \end{aligned}$$

En una $act_3 = \text{“classe”}$ d'un conjunt de membres $S \in SubCom$ on tracten temàtiques al voltant dels conceptes $C \subset Con$, podríem considerar un rol de suport, on els diàlegs $dlg \in Dlg$ consistirien en respostes com “Gràcies per participar!” o “És una pregunta molt interessant” quan els estudiants participessin. Les parts del diàleg memoritzades $M \in M_X$. Vist dins el marc que presentem:

$$\begin{aligned} &nom = \text{“Clara”}, personalitat = \text{“agraïda”}, modalitat = \text{“text”} \\ &Carac_D(dlg) = (C, \{act_3\}, M, S) \\ &(rol_4 = \text{“suport”}, \{nom, personalitat, modalitat\}) \in Rol_X(act_3), \\ &Acc_X(act_3, rol_4, M) = \{\text{“llençar missatge de suport”}\} \subset Acc \end{aligned}$$

En un $act_4 = \text{“projecte”}$ d'un grup $S \in SubCom$ abastant els conceptes $C \subset Con$ podria ser interessant un xatbot que prengué decisions de formació de grups, intentant equilibrar les agrupacions d'estudiants. A partir d'aquí podria fer un seguiment del funcionament del grup, mantenint diàlegs $dlg \in Dlg$ de forma individual o col·lectiva, preguntant com va el projecte (planificació de reunions, assistència a reunions, participació, etc.), amb parts del diàleg dlg memoritzades $M \in M_X$:

$$\begin{aligned} &nom = \text{“Xavi”}, personalitat = \text{“simpàtica”}, modalitat = \text{“text”}, plataforma = \\ &\text{“multiplataforma”}, Carac_D(dlg) = (C, \{act_4\}, M, S) \end{aligned}$$

$$(rol_5 = \text{“analitzador de grups”}, \{nom, personalitat, modalitat\}) \in Rol_X(act_4) \\ Acc_x(act_4, rol_5, M) = \{\text{“formar grup”}, \text{“fer seguiment”}\} \subset Acc$$

Per participar activament a una comunitat virtual és habitual haver de dur a terme un registre. Aquest registre es podria acompanyar d'un guia accessible a tots els membres que expliqués l'organització de la plataforma, mantenint diàlegs $dlg \in Dlg$ on la temàtica comprendria conceptes $C \subset Con$ bàsics d'entorns virtuals i educatius. Les parts del diàleg dlg memoritzades serien $M \in M_X$:

$$plataforma = \text{“monoplataforma”}, estil = \text{“formal”} \\ Carac_D(dlg) = (C, \text{“registre”}, M, Mem) \\ (rol_6 = \text{“acompanyant”}, \{plataforma, estil\}) \in Rol_X(\text{“registre”}), \\ Acc_X(\text{“registre”}, rol_6, M) = \{\text{“mostrar lloc web”}, \text{“explicar funcionament} \\ \text{comunitat”}, \text{“resumir històric d'activitats”}\}$$

3.2.2 Comunitats Virtuals de participació ciutadana

La participació ciutadana consisteix en la intervenció activa dels ciutadans i ciutadanes en els processos de presa de decisions públiques en funció dels interessos socials. Des de la perspectiva normativa, el terme de participació ciutadana pot restringir-se a aquells casos que representen una resposta, individual o col·lectiva, de la societat a una convocatòria realitzada per part de les autoritats governamentals. En la democràcia representativa vigent, en ocasions, una part de la societat és la que exerceix el poder de representació prenent decisions de forma legítima que afecten a una majoria.

En aquest apartat considerarem la plataforma digital “*decidim.barcelona*” [9] que va sorgir amb l'objectiu de promoure la participació ciutadana i la democràcia directa, deliberativa i participativa a la ciutat, combinant tant eines basades en les TICs com en recursos presencials.

Els ciutadans, i les persones interessades, poden participar més activament en actes, òrgans i processos participatius de la ciutat, donant l'opció a compartir les opinions polítiques a la plataforma. Totes elles poden registrar-se i publicar una iniciativa ciutadana, seguir un procés participatiu i les seves propostes, intervenir donant la opinió d'alguna proposta o comentari d'algun usuari. Veiem com encaixa la comunitat dins el marc que es proposa:

L'**ontologia** de la comunitat virtual comprendrà una diversitat de conceptes. La majoria d'ells giraran entorn política, ecologia, economia, sostenibilitat, urbanisme, o també privacitat i connectivitat. Per exemple, dins la plataforma trobem alguna definició de conceptes com “Els processos participatius són un seguit de trobades delimitades en un temps concret per promoure el debat i el contrast d'arguments entre la ciutadania, o entre aquesta i les persones responsables municipals” i “Els òrgans de participació són canals de trobada i d'interlocució regulars entre la ciutadania i l'Ajuntament per debatre i recollir opinions i propostes a fi d'incidir en les polítiques municipals”. Dins la formalització tindrem:

$con_1 =$ “processos participatius”, $con_2 =$ “òrgans de participació”

$def_1 =$ “seguit de trobades delimitades en un temps concret per promoure el debat i el contrast d’arguments entre la ciutadania, o entre aquesta i les persones responsables municipals”,

$def_2 =$ “canals de trobada i d’interlocució regulars entre la ciutadania i l’Ajuntament per debatre i recollir opinions i propostes a fi d’incidir en les polítiques municipals”

$Dic(con_1) = def_1, Dic(con_2) = def_2$

I considerant que els processos participatius es recolzen en els òrgans de participació:

$relc_1 =$ “es recolzen” $\in R_C, ConRel(con_1, con_2) = relc_1$

Veiem com és l’**estructura social** de la plataforma. Podem considerar el conjunt de membres Mem de tota la plataforma i $S_1, \dots, S_n \in SubCom$ com els subconjunts que corresponen als òrgans de participació de la ciutat o els subconjunts de membres que coincideixen fent activitats (involucrar-se en un procés participatiu, participar en un debat, intervenir en una proposta, etc.). Podríem, de nou, considerar les relacions del membres com:

$$Rel_M(m) \subseteq \bigcup_{\substack{1 \leq i \leq nI, \\ m \in S_i}} S_i, \quad m \in Mem$$

En quant a les propietats, els ciutadans que completen el registre poden donar dades com el nom i els interessos personals. Les subcomunitats poden tenir un lloc de trobada, un calendari de trobades o un objectiu, a més d’un nom. En termes de la formalització,

$$\begin{aligned} nom, interessos, ubicació, calendari, objectiu &\in Prop_P \\ \{nom, interessos\} &\subseteq Carac_P(m), m \in Mem, \\ \{nom, ubicació, calendari, objectiu\} &\subseteq Carac_P(s), s \in SubCom \end{aligned}$$

Els rols que sorgeixen entre els membres de la plataforma podrien correspondre’s al participant de processos participatius, seran la majoria de ciutadans qui desenvolupi aquest rol, els moderadors, apareixeran per garantir les bones dinàmiques als debats i el compliment de bones conductes a intervencions, l’informador, proporcionarà informació dels diferents processos participatius, i els avaluadors, professionals que es dedicaran a avaluar la viabilitat dels projectes. Per tant, queda:

$rol_1 =$ “participant”, $rol_2 =$ “moderadors”, $rol_3 =$ “informador”, $rol_4 =$ “avaluadors”, $rols_1, \dots, rols_4 \in Rols$

S’observa també l’**espai virtual** que podria definir la diversitat d’activitats que s’ofereixen. Tindríem en primer lloc el procés participatiu, una activitat informativa que connecta diverses activitats participatives com un debat o una votació, o

activitats discursives, com les propostes o les iniciatives, on tothom pot donar la seva opinió i, fins i tot, recolzar intervencions d'altres usuaris. El format en que es proporcionen aquestes activitats és híbrid: combinen la presencialitat per activitats més síncrones i participatives, amb la virtualitat per activitats informatives, discursives i asíncrones. La plataforma també permet als usuaris seguir processos o propostes que els interessin, respondre comentaris, posicionar-se a favor o en contra en una proposta o debat. Diversos exemples de la formalització d'activitats i accions serien:

$$act_1 = \text{"procés participatiu"}, act_2 = \text{"proposta"}, act_3 = \text{"debat"}, act_4 = \text{"votació"}, \\ act_1, \dots, act_4 \in Act$$

$$acc_1 = \text{"llegir contingut"}, acc_2 = \text{"baixar contingut"}, acc_3 = \text{"editar contingut"}, \\ acc_4 = \text{"pujar contingut"}, acc_5 = \text{"seguir"}, acc_6 = \text{"compartir opinió"}, \\ acc_7 = \text{"valorar viabilitat"}, acc_8 = \text{"intervenir"}, acc_9 = \text{"votar"}, acc_1, \dots, acc_9 \in Acc \\ nom, tipus, sincronia, caràcter, resultat, data_execució, presencial \in Prop_A$$

Si considerem $S_i \subset SubCom$ la subcomunitat que involucra tots els membres que participen a l'activitat act_i i $C_i \subset Con$ el subconjunt de conceptes que abasta, llavors una construcció de la caracterització de les activitats seria:

$$Carac_A(act_1 = \text{"procés participatiu"}) = (\{nom, tipus = \\ \text{"informativa participativa"}, sincronia = \text{"Fals"}, data_execució\}, S_1, C_1) \\ Carac_A(act_2 = \text{"proposta"}) = (\{nom, tipus = \text{"discursiva"}, resultat\}, S_2, C_2) \\ Carac_A(act_3 = \text{"debat"}) = (\{nom, tipus = \text{"discursiva participativa"}, sincronia = \\ \text{"Cert"}, presencial = \text{"Cert"}\}, S_3, C_3) \\ Carac_A(act_4 = \text{"votació"}) = (\{nom, tipus = \text{"participativa"}, caràcter = \\ \text{"individual"}, resultat, data_execució\}, S_4, C_4)$$

Per assignar els rols que podran prendre els diferents membres és necessària la concreció de la funció Rol_{AM} . Donat $m \in Mem$ veiem quins rols pot prendre i quines accions poden desenvolupar aquests rols:

- Per $act_1 = \text{"procés participatiu"}$ es podran desenvolupar dos rols: el de participant, que pot seguir el procés, llegir el contingut o descarregar-lo, així com seguir el procés, el d'informador, podrà llegir, baixar, pujar o editar el contingut.

$$Rol_{AM}(act_1, mem) \in \{rol_1, rol_3\} \\ Acc_{AR}(act_1, rol_1) = \{acc_1, acc_2, acc_5\}, \\ Acc_{AR}(act_1, rol_3) = \{acc_1, acc_2, acc_3, acc_4\}$$

- Per $act_2 = \text{"proposta"}$ es podran desenvolupar el rol de participant, el de moderador i el d'avaluador. El participant podrà seguir la proposta, llegir o baixar el contingut i compartir la seva opinió. El moderador podrà intervenir i l'avaluador podrà valorar la viabilitat.

$$\begin{aligned} Rol_{AM}(act_2, mem) &\in \{rol_1, rol_2, rol_4\} \\ Acc_{AR}(act_2, rol_1) &= \{acc_1, acc_2, acc_5, acc_6\}, \\ Acc_{AR}(act_2, rol_2) &= \{acc_8\}, Acc_{AR}(act_2, rol_4) = \{acc_7\} \end{aligned}$$

- Per act_3 = “debat” es podran desenvolupar el rol de participant i el de moderador. Serà similar en el cas anterior: el participant podrà compartir l’opinió i el moderador podrà intervenir.

$$\begin{aligned} Rol_{AM}(act_3, mem) &\in \{rol_1, rol_2\} \\ Acc_{AR}(act_3, rol_1) &= \{acc_6\} \\ Acc_{AR}(act_3, rol_2) &= \{acc_8\} \end{aligned}$$

- Per act_4 = “votació” es podrà desenvolupar de participant. Aquest, podrà simplement votar.

$$\begin{aligned} Rol_{AM}(act_4, mem) &\in \{rol_1\} \\ Acc_{AR}(act_4, rol_4) &= \{acc_9\} \end{aligned}$$

Les **convencions socials** d’aquesta comunitat tindran normes en comú amb altres pàgines web, com per exemple: $norm_1$ = “No es permet afegir cap contingut que infringeixi qualsevol dret de l’Ajuntament o de qualsevol tercera part, com ara el copyright, les marques registrades o altres drets de propietat intel·lectual i industrial o altres drets relacionats” $\in Norm$. També es contemplen d’altres que venen més especificades al tipus de pàgina web que és, com: $norm_2$ = “No es permet afegir cap contingut que faci campanyes promovent el vot massiu per a altres propostes no relacionades amb el procés i el marc de debat” $\in Norm$.

Per tal de completar una activitat és necessari seguir un protocol, per exemple és el cas en les diverses fases en que un procés participatiu es subdivideix. En cadascuna de les fases hi haurà involucrades altres activitats, com ara debats per discutir les propostes i finalment un procés de votació. En aquest sentit, considerem act_1 = “procés participatiu” i les activitats relacionades: $\{act_2^1, \dots, act_2^n\}$ les activitats del tipus act_2 = “proposta”, $\{act_3^1, \dots, act_3^m\}$ les activitats del tipus act_3 = “debat” i $\{act_4^1, \dots, act_4^k\}$ les del tipus act_4 = “votació”. Llavors:

$prot_1$ = “presentació i informació del procés participatiu, fase de diagnosi, on es donen debats per contrastar diverses propostes, fase de votació i aprovació, i fase de retorn, on recopilen la síntesi del procés, amb l’explicació de l’execució de les propostes” $\in Prot$,

$$ActRel(act_1) = \{act_2^1, \dots, act_2^n, act_3^1, \dots, act_3^m, act_4^1, \dots, act_4^k\}$$

Dins la comunitat virtual de **decidim.barcelona** també podríem construir un **xat-bot** que pogués fer un anàlisi del perfil d’usuaris, propostes i intervencions per tal d’interactuar amb ells i impulsar noves interaccions a la plataforma, millorant l’experiència de la persona usuària. Per exemple:

Es poden crear perfils d'usuari guardant les interaccions que realitza cada persona usuària en la comunitat (propostes més seguides, les intervencions que fa, etc.). A partir d'aquí, un xatbot podria beneficiar-se d'aquesta informació per poder recomanar altres propostes, demanar l'opinió d'una proposta que hagi causat repercussió, informar de l'estat de votació de propostes suggerides, veure la repercussió que té amb usuaris d'un perfil similar, etc.

Ell xatbot comptaria amb un sèrie de propietats assignables a les activitats en les que es pot involucrar, amb les parts del diàlegs dlg memoritzades $M \in M_X$ que manté ab les diferents persones participants:

$$\begin{aligned} nom &= \text{"decidim.bot"}, gènere = \text{"no binari"}, plataforma = \text{"monoplataforma"} \\ nom, gènere, plataforma &\in Prop_X \end{aligned}$$

Els diàlegs que mantendria amb les persones participants tindrien una relació directa amb els conceptes $C \subset Con$ que involucressin les propostes $A \subset Act$. En una $act_2 = \text{"proposta"} \in A$ el xatbot funcionaria com un recomanador, amb una personalitat amigable i que es moguéss dins la plataforma. També, donat el rol, el xatbot podrà dur a terme certes accions com recomanar una proposta i a l'hora que faciliti l'accés a compartir l'opinió:

$$\begin{aligned} personalitat &= \text{"amigable"}, modalitat = \text{"text"} \in Prop_X, \\ Carac_D(dlg) &= (C, A, M, Mem) \\ (rol_1 = \text{"recomanador"}, \{nom, gènere, plataforma, personalitat, modalitat\}) &\in \\ &Rol_X(act_2) \\ Acc_X(act_2, rol_1, M) &= \{\text{recomanar proposta, demanar opinió}\} \cup Acc_1 \subset Acc \end{aligned}$$

Un altre escenari on el xatbot pot beneficiar a la participació és en d'incloure'l com informador de act_4 votacions actives que potencialment poden atraure la persona usuària. Comptaria amb una personalitat neutre amb un estil de llenguatge formal

$$\begin{aligned} personalitat &= \text{"neutre"}, estil = \text{"formal"} \in Prop_X, \\ Carac_D(dlg) &= (C, A, M, Mem) \\ (rol_2 = \text{"informador"}, \{nom, gènere, plataforma, personalitat, estil\}) &\in \\ &Rol_X(act_4) \\ Acc_X(act_4, rol_2, M) &= \{\text{"informar votació activa"}\} \cup Acc_1 \subset Acc \end{aligned}$$

Amb l'objectiu de continuar amb aquesta línia de treball, on un xatbot aprofita la memòria de diàlegs per la recollida de dades, al següent capítol es presenta un model aplicable de la plataforma de `decidim.barcelona` amb el qual es podran obtenir resultats de consultes per la comunitat virtual del tipus: "Donat el perfil d'una persona usuària, quines propostes en votació s'alineen més amb els seus interessos?".

Capítol 4

Model de comunitat virtual `decidim.barcelona`

En aquest apartat, s'explicarà el marc de treball dels models gràfics probabilístics (o PGM, per les sigles en anglès) i, posteriorment, es passarà la proposta d'un model aplicable de la comunitat virtual `decidim.barcelona`. Es farà un anàlisi de la pàgina web i la seva API per definir una sèrie de variables que ocupin una rellevància a la incertesa que es vulgui resoldre.

La pròxima secció tractarà sobre models gràfics probabilístics, la teoria de la qual ha estat obtinguda a [16].

4.1 Models Gràfics Probabilístics

Els sistemes complexos es caracteritzen per la presència de múltiples aspectes interrelacionats, molts d'ells connectats amb un raonament. Per exemple, realitzant un diagnòstic mèdic són múltiples les malalties que pot tenir un pacient. Existeixen desenes o centenars de símptomes i proves diagnòstiques, característiques personals i factors genètics que el predisposen a la malaltia, i moltes més consideracions a tenir en compte. Aquests dominis es poden caracteritzar en termes d'un conjunt de variables aleatòries, on el valor de cadascuna de les variables defineix una propietat important del món. Per exemple, una malaltia concreta, com la grip, pot ser una variable del nostre domini que pren dos valors, present o absent. Un símptoma, com la febre, pot ser una variable en el nostre domini que pren valors continus. El conjunt de possibles variables i els seus valors és una decisió important de disseny i depèn molt de les preguntes que volem respondre sobre el domini.

L'objectiu de desenvolupar un PGM consisteix en poder raonar probabilísticament sobre els valors d'una o més variables del model, possiblement donades observacions sobre algunes altres variables. Per fer-ho utilitzant un raonament probabilístic basat en principis, s'ha de construir una distribució conjunta sobre l'espai de possibles assignacions a algun conjunt de variables aleatòries X . Aquest tipus de model permet respondre un ampli ventall de consultes interessants. Per exemple, podem fer

l'observació que una variable X_i pren el valor específic x_i i preguntar, en la posterior distribució resultant, quin és el valor més probable d'una altra variable X_j , és a dir,

$$\arg \max_{x_j \in \text{Val}(X_j)} P(X_j = x_j \mid X_i = x_i)$$

4.1.1 Exemple

Suposem que ens interessa un entorn de diagnòstic mèdic molt senzill on ens centrem en dues malalties: la grip i l'al·lèrgia al pol·len. Aquestes no s'exclouen mútuament, ja que un pacient pot tenir qualsevol, ambdues o cap. Així, podríem tenir dues variables aleatòries de valors binaris, la grip G i l'al·lèrgia al pol·len A . També tenim una variable aleatòria estació E de 4 valors, que es correlaciona tant amb la grip com amb l'al·lèrgia. També podem considerar tenir dos símptomes, la congestió C i el dolor muscular D , cadascun dels quals també té valors binaris. En general, l'espai de probabilitat té $2 \times 2 \times 4 \times 2 \times 2 = 64$ valors, corresponents a les possibles assignacions a aquestes cinc variables. Una manera de representar el sistema seria la figura 4.1

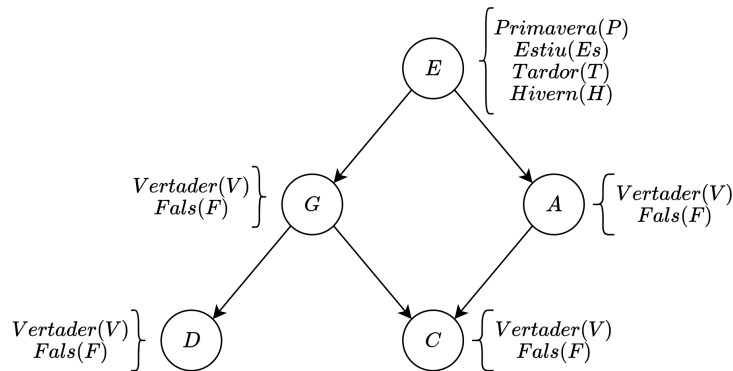


Figura 4.1: En aquesta representació gràfica els nodes corresponen a les variables amb els seus respectius dominis i les arestes corresponen a interaccions probabilístiques directes entre elles.

Donada una distribució conjunta en aquest espai, podem, per exemple, fer preguntes com la probabilitat que el pacient tingui la grip ja que és tardor, i que tingui congestió, però sense dolor muscular; com a expressió de probabilitat condicionada, aquesta consulta s'indicaria:

$$P(G = V \mid E = T, C = V, D = F)$$

Hi ha una perspectiva dual que es pot utilitzar per interpretar l'estructura del graf 4.1. Des d'una banda, el graf és una representació compacta d'un conjunt d'independències que es mantenen en la distribució. Aquestes propietats prenen la forma 'X és independent de Y donada Z', denotada com a $(X \perp Y \mid Z)$, per a

alguns subconjunts de variables X, Y, Z . Seguint l'exemple anterior, se satisfà la independència condicional

$$(C \perp E \mid G, A).$$

És a dir, si interessa saber si un pacient té congestió i sabem que té grip i al·lèrgia, llavors l'estació ja no és informativa. Aquesta afirmació no implica que l'estació sigui independent de la congestió; només que tota la informació que podem obtenir de l'estació sobre les possibilitats de tenir congestió ja l'obtenim en saber si el pacient té grip i té al·lèrgia.

L'altra perspectiva és que el graf defineix un esquelet per representar de manera compacta una distribució dimensional alta: en lloc de codificar la probabilitat de cada assignació possible a totes les variables del nostre domini, podem “descompondre” la distribució en factors més petits, cadascun en un espai de possibilitats molt més reduït. Aleshores, podem definir la distribució conjunta global com a producte d'aquests factors, l'anomenada *regla de la cadena per xarxes Bayesianes*. La factorització de la distribució associada al graf 4.1 seria

$$\begin{aligned} P(E = e, G = g, A = a, D = d, C = c) &= P(E = e) \cdot P(G = g \mid E = e) \cdot \\ &\cdot P(A = a \mid E = e) \cdot P(D = d \mid G = g) \cdot \\ &\cdot P(C = c \mid G = g, A = a) \end{aligned}$$

on $e \in \text{Val}(E), g \in \text{Val}(G), a \in \text{Val}(A), d \in \text{Val}(D), c \in \text{Val}(C)$, on utilitzem $\text{Val}(X)$ per indicar el conjunt de valors que pot prendre una variable aleatòria X .

Donada una distribució com l'anterior, parlem de distribució de probabilitat condicional (o CPD, per les segles en anglès) sobre una variable quan aquesta depèn d'altres, com és el cas de *Grip* o de *Congestió*.

4.1.2 Xarxes Bayesianes

Després de les intuïcions desenvolupades a l'exemple anterior, veiem la definició formal de la semàntica d'una estructura de xarxa Bayesiana. Per una banda, el graf representava un conjunt d'independències entre variables del tipus $(X \perp Y \mid Z)$.

Definició 4.1.1. L'estructura d'una xarxa Bayesiana \mathcal{G} és un graf acíclic i dirigit els nodes del qual representen variables aleatòries X_1, \dots, X_n . Denotem $Pa_{X_i}^{\mathcal{G}}$ els nodes pares de X_i a \mathcal{G} i $NoDesc_{X_i}$ les variables del graf \mathcal{G} que no són descendents de X_i . Llavors \mathcal{G} codifica el següent conjunt de suposicions d'independència condicionada, anomenades independències locals, i denotades per $\mathcal{I}_\ell(\mathcal{G})$:

$$\text{Per cada variable } X_i : (X_i \perp NoDesc_{X_i} \mid Pa_{X_i}^{\mathcal{G}}).$$

En altres paraules, les independències locals afirmen que cada node X_i és condicionalment independent als seus no descendents donats els seus pares.

Per altra banda, l'exemple era un graf anotat amb CPDs, que definia una distribució conjunta mitjançant la regla de la cadena per xarxes Bayesianes. Veurem

que aquestes dues definicions són, de fet, equivalents Una distribució P satisfà les independències locals associades al graf \mathcal{G} si i només si P és representable com un conjunt de CPDs associat amb el graf \mathcal{G} .

Es considera inicialment la definició d'independència condicional i algunes notacions que s'utilitzaran:

Definició 4.1.2. Siguin X, Y, Z conjunts de variables aleatòries. Direm que X és condicionalment independent de Y donat Z a la distribució P si P satisfà

$$(X \perp Y \mid Z),$$

és a dir,

$$P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z) \cdot P(Y = y \mid Z = z), \\ \forall x \in Val(X), \forall y \in Val(Y), \forall z \in Val(Z),$$

Ens referirem a les variables al conjunt Z com les variables observades.

Exemple 4.1.3. Podem veure les independències que compleix l'estructura de xarxa Bayesiana que representa la figura 4.1:

$$(G \perp A \mid E), \\ (A \perp G, D \mid E) \\ (D \perp E, A, C \mid G), \\ (C \perp E, D \mid G, A)$$

A partir de les propietats de descomposició i simetria de la independència condicional s'obtenen totes les independències que il·lustra el graf.

Notació 1. Donada una distribució P sobre les variables X_1, \dots, X_n que pren valors $x_i \in Val(X_i)$. S'escriurà $P(x_1, \dots, x_n)$ per referir-nos a $P(X_1 = x_1, \dots, X_n = x_n)$. La notació que se seguirà per iterar sobre el conjunt de valors d'una variable aleatòria X és:

$$\sum_{x_i} P(x_i) = \sum_{x_i \in Val(X)} P(X = x_i)$$

que en aquest cas val 1.

Considerarem ara la definició del conjunt d'independències associades a una distribució P :

Definició 4.1.4. Sigui P una distribució de probabilitat sobre un conjunt de variables aleatòries \mathcal{X} . Definim $\mathcal{I}(P)$ com el conjunt d'expressions de la forma $(X \perp Y \mid Z)$ que conté P , on $X, Y, Z \subset \mathcal{X}$.

Per tal de reescriure l'afirmació “ P satisfà les independències locals associades a \mathcal{G} ” de manera que relacioni els conjunts d'independència s'introdueix el concepte d' \mathcal{I} – mapa (mapa d'independències):

Definició 4.1.5. Sigui \mathcal{G} una estructura de xarxa Bayesiana. Direm que \mathcal{G} és un \mathcal{I} – mapa per a un conjunt d'independències \mathcal{I} si $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}$.

Notació 2. Direm que \mathcal{G} és un \mathcal{I} – mapa per P si \mathcal{G} és un \mathcal{I} – mapa per $\mathcal{I}(P)$.

Finalment, el graf representa una distribució conjunta d'alta dimensió, que se simplifica descomponent en factors d'espais de possibilitats més reduïts.

Definició 4.1.6. Sigui \mathcal{G} un graf amb estructura de xarxa Bayesiana sobre les variables X_1, \dots, X_n . Direm que una distribució conjunta P , sobre el mateix espai, factoritza d'acord amb \mathcal{G} si P es pot expressar com el producte:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid Pa_{X_i}^{\mathcal{G}})$$

Es coneix com la regla de la cadena per xarxes Bayesianes. Els factors individuals $P(X_i \mid Pa_{X_i}^{\mathcal{G}})$ s'anomenen distribucions locals de probabilitat en cas que $Pa_{X_i}^{\mathcal{G}} = \emptyset$, o distribucions de probabilitat condicionada (CPDs), altrament.

Podem donar ja la definició de xarxa Bayesiana, on es contemplen les dues perspectives vistes d'un graf com el tractat a l'exemple de l'apartat anterior.

Definició 4.1.7. Una xarxa Bayesiana és el parell $\mathcal{B} = (\mathcal{G}, P)$ on P factoritza sobre \mathcal{G} .

Amb l'objectiu de provar que “una distribució P satisfà les independències locals associades al graf \mathcal{G} si i només si P és representable com un conjunt de CPDs associat amb el graf \mathcal{G} ” es presenten els següents dos teoremes:

Teorema 4.1.8. Sigui \mathcal{G} una estructura de xarxa Bayesiana sobre un conjunt de variables aleatòries \mathcal{X} , i sigui P una distribució conjunta sobre el mateix espai. Si \mathcal{G} és un \mathcal{I} – mapa per P , llavors P factoritza sobre \mathcal{G} .

Demostració. Suposem, sense pèrdua de generalitat, que X_1, \dots, X_n és un ordre topològic de les variables de \mathcal{X} relatiu a \mathcal{G} . Utilitzem la regla de la cadena per probabilitats:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid X_1, \dots, X_{i-1}).$$

Ara, considerem un dels factors $P(X_i \mid X_1, \dots, X_{i-1})$. Com que \mathcal{G} és un \mathcal{I} – mapa per P , tenim $(X_i \perp NoDesc_{X_i} \mid Pa_{X_i}^{\mathcal{G}}) \in \mathcal{I}(P)$. Per suposició, tots els nodes pares de la variable X_i estan en el conjunt X_1, \dots, X_{i-1} . A més, cap node descendent de X_i pertany a aquest conjunt. Per tant,

$$\{X_1, \dots, X_{i-1}\} = Pa_{X_i}^{\mathcal{G}} \cup Z,$$

on $Z \subseteq NoDesc_{X_i}$. De les independències locals d' X_i i de la propietat de descomposició s'obté que $(X_i \perp Z \mid Pa_{X_i}^{\mathcal{G}})$. Per tant, tenim

$$P(X_i \mid X_1, \dots, X_{i-1}) = P(X_i \mid Pa_{X_i}^{\mathcal{G}}).$$

Aplicant la transformació a tots els factors de la descomposició de la regla de la cadena obtenim que P factoritza sobre \mathcal{G} , tal com volíem. \square

Així, les suposicions d'independència condicional induïdes per una estructura de xarxa Bayesiana \mathcal{G} ens permeten factoritzar una distribució P , per a la qual \mathcal{G} és un \mathcal{I} – mapa, en petites CPDs.

Veiem ara la implicació contrària:

Teorema 4.1.9. Sigui \mathcal{G} una estructura de xarxa Bayesiana sobre un conjunt de variables aleatòries \mathcal{X} i, sigui P una distribució conjunta sobre el mateix espai. Si P factoritza sobre \mathcal{G} , aleshores \mathcal{G} és un \mathcal{I} – mapa per P .

Demostració. Suposem, sense pèrdua de generalitat, que X_1, \dots, X_n és un ordre topològic de les variables de \mathcal{X} relatiu a \mathcal{G} . Per veure la demostració d'aquest teorema requerirem d'un lema previ:

Lema 4.1.10. Si $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid Pa_{X_i}^{\mathcal{G}})$, aleshores $P(X_1, \dots, X_k) = \prod_{i=1}^k P(X_i \mid Pa_{X_i}^{\mathcal{G}})$ per $\forall k \leq n$.

Demostració. Per $k = n$ obtenim la suposició. Prenem $k < n$. Considerem la distribució conjunta sobre les variables X_1, \dots, X_k , la qual equival a considerar la distribució conjunta sobre X_1, \dots, X_n observant tots els esdeveniments X_{k+1}, \dots, X_n :

$$P(X_1, \dots, X_k) = \sum_{x_n} \dots \sum_{x_{k+1}} P(X_1, \dots, X_k, x_{k+1}, \dots, x_n)$$

Ara, la distribució es pot expressar mitjançant la factorització de P sobre \mathcal{G} .

$$P(X_1, \dots, X_k) = \sum_{x_n} \dots \sum_{x_{k+1}} \prod_{i=1}^n P(X_i \mid Pa_{X_i}^{\mathcal{G}})$$

L'observació de les variables X_{k+1}, \dots, X_n no afecta la distribució de $P(X_i \mid Pa_{X_i}^{\mathcal{G}})$, ja que, tal com estan ordenades les variables, $X_j \notin Pa_{X_i}^{\mathcal{G}}$ per tot $1 \leq i \leq k, k+1 \leq j \leq n$. Llavors obtindrem factors $\sum_{x_j} P(x_j \mid Pa_{X_j}^{\mathcal{G}}) = 1$, per $k+1 \leq j \leq n$:

$$P(X_1, \dots, X_k) = \sum_{x_n} \dots \sum_{x_{k+1}} \left(\prod_{i=1}^{n-1} P(X_i \mid Pa_{X_i}^{\mathcal{G}}) \cdot P(x_n \mid Pa_{X_n}^{\mathcal{G}}) \right)$$

Com que X_1, \dots, X_n està ordenat topològicament, no existeix cap $i < j$ tal que

$X_j \in Pa_{X_i}^{\mathcal{G}}$. Per tant, ho podem reescriure com

$$\begin{aligned}
P(X_1, \dots, X_k) &= \sum_{x_{n-1}} \dots \sum_{x_{k+1}} \left(\prod_{i=1}^{n-1} P(X_i | Pa_{X_i}^{\mathcal{G}}) \right) \cdot \left(\sum_{x_n} P(x_n | Pa_{X_n}^{\mathcal{G}}) \right) \\
&= \sum_{x_{n-1}} \dots \sum_{x_{k+1}} \left(\prod_{i=1}^{n-1} P(X_i | Pa_{X_i}^{\mathcal{G}}) \right) \cdot 1 \\
&= \sum_{x_{n-2}} \dots \sum_{x_{k+1}} \left(\prod_{i=1}^{n-2} P(X_i | Pa_{X_i}^{\mathcal{G}}) \right) \cdot \left(\sum_{x_{n-1}} P(x_{n-1} | Pa_{X_{n-1}}^{\mathcal{G}}) \right) \\
&\quad \vdots \\
&= \prod_{i=1}^k P(X_i | Pa_{X_i}^{\mathcal{G}})
\end{aligned}$$

□

Seguint amb la demostració del teorema, s'ha de veure que donat $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa_{X_i}^{\mathcal{G}})$, es té $\mathcal{I}_\ell(\mathcal{G}) \subseteq \mathcal{I}(P)$. Això, és equivalent a veure que $\forall i (X_i \perp NoDesc_{X_i} | Pa_{X_i}^{\mathcal{G}}) \in \mathcal{I}(P)$.

Observem primerament que $\{X_1, \dots, X_i\} \subseteq \{X_i, NoDesc_{X_i}\}$, per tant, podem considerar $\{X_i, NoDesc_{X_i}\} = \{Y_1, \dots, Y_m\}$ amb $m \geq i$ i on

$$Y_j = \begin{cases} X_j & j - \text{èsim } NoDesc_{X_i}, \\ X_i & j = m \end{cases} \quad (4.1.1)$$

Observem que $\{Y_1, \dots, Y_m\}$ mantenen tant l'ordre topològic com la factorització sobre \mathcal{G} :

$$P(X_1, \dots, X_n) = P(Y_1, \dots, Y_n) = \prod_{i=1}^n P(Y_i | Pa_{Y_i}^{\mathcal{G}})$$

Així, pel lema anterior

$$\begin{aligned}
P(X_i | NoDesc_{X_i}) &= \frac{P(X_i, NoDesc_{X_i})}{P(NoDesc_{X_i})} \\
&= \frac{P(Y_1, \dots, Y_m)}{P(Y_1, \dots, Y_{m-1})} \\
&= \frac{\prod_{j=1}^m P(Y_j | Pa_{Y_j}^G)}{\prod_{j=1}^{m-1} P(Y_j | Pa_{Y_j}^G)} \\
&= P(Y_m | Pa_{Y_m}^G) \\
&= P(X_i | Pa_{X_i}^G)
\end{aligned}$$

La qual cosa significa que, per cada i , $(X_i \perp NoDesc_{X_i} | Pa_{X_i}^G) \in \mathcal{I}(P)$ □

4.2 Model de la comunitat virtual decidim.barcelona

La construcció de xarxes Bayesianes se centra en el procés de passar d'una distribució determinada a una xarxa Bayesiana. Tanmateix, la vida real no és així. Tenim un model imprecís del món i hem de cristal·litzar-lo en una xarxa amb estructura i paràmetres. Aquesta tasca es divideix en diverses components, cadascuna de les quals pot ser força subtil, ja que els errors de modelització poden tenir conseqüències importants per a la qualitat de les respostes obtingudes de la xarxa, o per al cost d'utilitzar la xarxa a la pràctica.

En aquest apartat, prendrem la comunitat virtual de `decidim.barcelona`. En concret, comprendrem quins són en els recursos que hi ha a l'abast de l'usuari a la pàgina web i veurem com obtenir un model a partir del qual un xatbot pugui ser capaç de treure'n profit.

4.2.1 Variables

Quan modelem un domini, hi ha moltes maneres possibles de descriure les entitats rellevants i els seus atributs. Escollir quines variables aleatòries utilitzar en el model és sovint una de les tasques més difícils, i aquesta decisió té implicacions en tot el model.

En termes generals, volem que el model contingui variables que potencialment puguem observar i d'altres que vulguem consultar, variables no observades. De vegades és interessant agregar una variable no observada que no interessa directament, i d'altres, simplement, es deixen d'afegir per la seva irrellevància dins el context.

També és important especificar un domini raonable de valors per a les nostres variables. En particular, si la nostra partició no és prou bona, els supòsits d'independència condicional poden ser falsos.

A partir d'aquí, veiem quines poden ser les variables que donin forma al model.



Figura 4.2: Diferents propostes recollides de la pàgina web de decidim.barcelona

Variables observades

Fixant-nos en la pàgina web i la seva API ens hem centrat en les interaccions visibles que es donen als processos participatius, més concretament, les propostes que es recullen dels usuaris i altres organitzacions. El funcionament d'un procés participatiu normalment consisteix en la formació de grups de debat amb un mateix interès, sorgint propostes que eventualment es penjen a la plataforma. En total, la plataforma compte amb un ventall de més de 27.000 propostes que serviran per fer inferència sobre el model.

Donant una primera ullada a les propostes ens trobem amb diferents colors (figura 4.2). El verd indica que la proposta ha estat acceptada i, si pertany a una selecció de candidats a la seva execució, aleshores entrarà en un procés de votació. Si la proposta és vermella, la proposta haurà estat denegada i no entrarà a la votació. Existeix també el color groc, indica que la proposta està en revisió. Finalment, hi ha el color blanc, que fa referència a aquelles propostes que compten amb sessions informatives.

Són en les propostes on les persones participants poden comentar 4.4a, començar a seguir o mostrar interès 4.4b, a més de poder respondre altres persones usuàries o mostrar el seu acord o desacord amb elles, com a 4.3.

Veiem que totes elles tenen en comú una sèrie de variables: la quantitat de seguidors, l'interès i els comentaris. De la seva API (figura 4.7) podem extreure les dades que corresponen a la zona de participació i les categories, que engloben l'àmbit d'actuació de les propostes. També hi ha accés als



Figura 4.3: Comentaris extrets d'una proposta. En ella s'observa un comentari en contra, amb 7 acords i 1 desacord, i un altre a favor, amb 1 acord i 5 desacords.

comentaris (veure la figura 4.3), obtenint de cadascun l'alineació amb la proposta, és a dir, si el comentari està a favor, en contra o és neutre respecte la proposta, i el suport d'altres usuaris (mostrat com a nombre d'acords \wedge i desacords \vee).

De totes les variables que s'han observat, considerarem per al nostre model les variables aleatòries següents:

1. Cada proposta p vindrà caracteritzada per:
 - (1) Les zones Z^p de participació que abasti.
 - (2) Les categories C^p associades.
 - (3) Una persona usuària té l'opció de “seguir” (4.4b) una proposta. No obstant, de cada proposta p només s'observa el recompte de seguidors S^p .
 - (4) Quan s'obre el període de votació d'una proposta, totes les persones usuàries poden votar a favor de l'execució de la proposta o abstenir-se. El recompte de vots R^p que rep la proposta serà un valor observat.
2. Per cada persona usuària u i cada proposta p :
 - (1) L'alineació A_u^p , que correspon a si l'usuari està a favor, recolzant la proposta per portar-la a l'execució, o en contra. La considerarem una variable parcialment observada, partint de la idea que els usuaris poden decidir no mostrar l'alineació al comentari.
 - (2) L'interès I_u^p que tingui l'usuari. Cada persona participant que navega per les propostes, pot decidir interessar-se mitjançant l'ús del “m'agrada” 4.4b. La considerarem una variable parcialment observada donat que pot haver propostes no visualitzades que puguin ser de l'interès de la persona usuària.
3. Donada una intervenció z a la proposta p de la persona usuària u , altres persones usuàries poden mostrar l'acord o el desacord amb aquest. La repercussió $R_{z,u}^p$ serà la variable que representi aquesta interacció.



(a) Acció de comentar, on la persona usuària pot posicionar-se a favor de la proposta, en contra o neutre.

(b) Acció de seguir i mostrar interès. La principal diferència resideix en que només se sap la quantitat de seguidors, mentre que sabem qui mostra interès amb el “M'agrada”

Figura 4.4: Accions que poden prendre les persones usuàries amb un rol de participant

The image shows a GraphQL IDE interface with a query on the left and its JSON response on the right. The query is a nested structure for 'participatoryProcesses' with various fields like 'components', 'proposals', 'nodes', 'category', 'scope', 'endorsements', 'state', '#voteCount', and '#comments'. The response is a JSON object with a 'data' field containing an array of 'participatoryProcesses' objects. Each object in the array has a 'title' (with a 'translation' field), 'category', 'scope', 'endorsements', and 'state'.

```

1 {
2   participatoryProcesses {
3     components {
4       ... on Proposals {
5         proposals {
6           nodes {
7             title{translation(locale:"ca")}
8             category {
9               name{translation(locale:"ca")}
10            }
11            scope {
12              name{translation(locale:"ca")}
13            }
14            endorsements {
15              id
16            }
17            state
18            #voteCount
19            #comments {
20              #author {
21                #id
22              }
23              #alignment
24              #upVotes
25              #downVotes
26              #comments {
27                #author {
28                  #id
29                }
30                #downVotes
31                #upVotes
32              }
33            }
34          }
35        }
36      }
37    }
38  }
39 }

```

```

{
  "data": {
    "participatoryProcesses": [
      {
        "components": [
          {
            "proposals": {
              "nodes": [
                {
                  "title": {
                    "translation": "Sensibilitzar sobre la
diversitat funcional"
                  },
                  "category": {
                    "name": {
                      "translation": "Autonomia personal "
                    }
                  },
                  "scope": {
                    "name": {
                      "translation": "Gràcia"
                    }
                  },
                  "endorsements": [],
                  "state": "accepted"
                },
                {
                  "title": {
                    "translation": "Festes majors
inclusives"
                  },
                  "category": {
                    "name": {
                      "translation": "Autonomia personal "
                    }
                  },
                  "scope": {
                    "name": {
                      "translation": "Gràcia"
                    }
                  },
                  "endorsements": [],
                  "state": "accepted"
                },
                {
                  "title": {
                    "translation": "Posar en funcionament
l'EB L'Univers amb gestió pública"
                  },
                  "category": {
                    "name": {

```

Figura 4.5: Consulta a l'API del web decidim.barcelona on podem extreure totes les dades.

Variables no observades

Per altra banda, també tindrem en consideració altres variables que no observem directament, però que poden influir dins el model. Seran potencialment les variables que el xatbot pugui consultar per extreure'n conclusions i poder actuar per tal que hi hagi més interacció dins la plataforma. Les que hem considerat són:

1. Per cada proposta p , un perfil de proposta P_p , que modeli la proposta.
2. Per cada usuari u :
 - (1) Un perfil d'usuari P_u que modeli l'usuari.
 - (2) Un zona Z_u des de la qual l'usuari pugui participar activament votant a una proposta.
 - (3) Una capacitat d'influència C_u que influirà sobre la repercussió, tant positiva com negativa, de les intervencions que faci l'usuari u .
3. Per cada proposta acceptada p i un usuari u que pugui participar, un vot V_u^p . Serà considerat una variable no observada ja que tota la informació que s'obté és el recompte de vots, en cap cas se sap quins usuaris u han votat per l'execució de la proposta p .

4.2.2 Estructura i probabilitats

Hi ha moltes estructures que són coherents amb el mateix conjunt d'independències de les nostres variables així com distribucions de probabilitat condicionada que compleixen les independències. Un enfocament que sol funcionar és triar una estructura que reflecteixi l'ordre causal i les dependències, de manera que les causes siguin pares de l'efecte. La causalitat l'observem del món, però no en el nostre procés d'inferència. Obtenir les distribucions de les variables és una mica més fàcil en el context dels models causals, ja que els paràmetres solen ser naturals i més interpretables.

A l'hora de determinar l'estructura, però s'ha de tenir en compte que les aproximacions són inevitables. Per a molts parells de variables, podem construir un escenari on una depengui de l'altra. Tot i que les estimacions de les probabilitats no han de ser exactes, hi ha altres errors que s'eviten, ja que poden tenir un efecte significatiu:

- Assignar una probabilitat 0 a un succés molt poc probable, però no impossible. El problema és que mai no es pot condicionar una probabilitat zero, per més proves que tinguem. Quan un esdeveniment és improbable, però no impossible, donar-li probabilitat zero es garantirà que conduirà a errors irrecuperables.
- Ordres de magnitud: petites diferències en esdeveniments de molt baixa probabilitat poden fer una gran diferència a les conclusions de la xarxa.

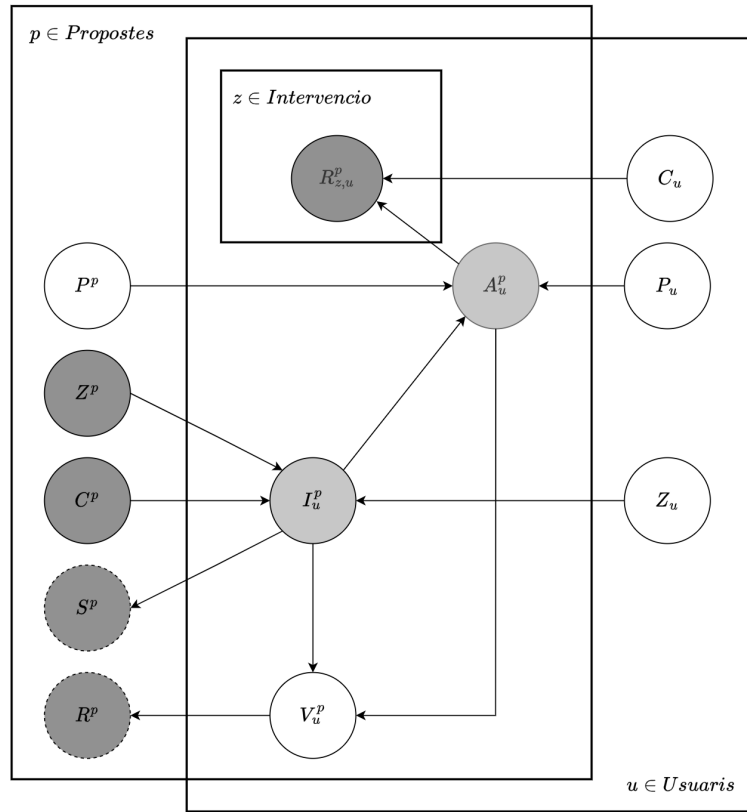


Figura 4.6: Estructura de xarxa Bayesiana del model proposat. Les variables observades i parcialment observades prenen un color més fosc. Les variables deterministes tenen el perímetre dels nodes puntejat.

- **Valors relatius:** el comportament qualitatiu de les conclusions a les quals arriba la xarxa, és a dir, el valor que té la probabilitat més alta, és bastant sensible a les mides relatives de $P(x | y)$ per a diferents valors y de Pa_X^G .

De les variables plantejades a l'apartat anterior, l'estructura triada és mostrada a la figura 4.6. Cada requadre exterior indica que les variables que agrupa es repeteixen per cada element del conjunt. És a l'estructura on es poden observar les diferents dependències i independències que reflecteixen l'ordre causal que s'ha pogut deduir. Les distribucions també veuen reflectides aquestes dependències:

- El perfil de la proposta P^p i el d'usuari P_u seguiran una distribució normal amb esperança i variància parametritzades,

$$P^p \sim \mathcal{N}(\mu_1, 1/\eta_1), \quad P_u \sim \mathcal{N}(\mu_2, 1/\eta_2)$$

Serà mitjançant la inferència que aquests paràmetres prendran valors i podrem deduir-ne resultats.

- Les zones Z^p , Z_u i la categoria C^p són variables categòriques, és a dir, si l'espai mostral de Z^p i Z_u és $\Omega_Z = \{z_1, \dots, z_n\}$ i $\Omega_C = \{c_1, \dots, c_m\}$ és el de C^p , llavors

a cada possible resultat li correspon la seva probabilitat,

$$P(Z^p = z_i) = \alpha_i, \quad 1 \leq i \leq n-1, \quad P(Z^p = z_n) = 1 - \sum_{i=1}^{n-1} \alpha_i$$

$$P(Z_u = z_i) = \beta_i, \quad 1 \leq i \leq n-1, \quad P(Z_u = z_n) = 1 - \sum_{i=1}^{n-1} \beta_i$$

$$P(C^p = c_i) = \xi_i, \quad 1 \leq i \leq m-1, \quad P(C^p = c_m) = 1 - \sum_{i=1}^{m-1} \xi_i$$

Cada cop que s'observa interacció per part de la persona usuària u a la proposta p amb zona z_i i categoria c_i , els valors d' α_i , β_i i ξ_i varien, de manera que es poden aproximar zones i categories d'interès.

- A l'anterior apartat s'observava la possibilitat que tenen les persones usuàries de mostrar interès per una proposta. Definíem una variable aleatòria per tal de modelar aquest interès. En aquest cas, s'ha considerat l'ordre causal, de manera que sigui la coincidència de zones i la categoria que tracti la proposta les que influeixin en l'interès I_u^p . El valor que pugui prendre l'interès serà binari, si la persona usuària està interessada ($I_u^p = 1$) o no ($I_u^p = 0$), per tant, serà una variable que segueixi una Bernoulli. El paràmetre serà més gran (i, en conseqüència, la probabilitat que tingui interès) quan hi hagi coincidència de zones entre Z^p i Z_u , i quan les categories associades tinguin paràmetres més grans (i.e. siguin d'interès):

$$I_u^p \sim \text{Bernoulli}(p_3), \quad p_3 = P(Z^p = z) * P(Z_u = z) + P(C^p = c) - \sigma_1, \quad \sigma_1 \in [-2, 1]$$

- A l'anterior apartat s'observava l'alineació en alguns comentaris de persones usuàries a les propostes. Suposarem que totes les persones participants u generen una alineació A_u^p de la proposta p , encara que puguin no compartir-la. El valor que pugui prendre aquesta variable aleatòria serà binari, a favor ($A_u^p = 1$) o en contra ($A_u^p = 0$), per tant, serà una variable que segueixi una Bernoulli. El paràmetre d'aquesta vindrà determinat pels perfils de proposta i usuari i , també, per l'interès de la persona participant en la proposta. Quan els perfils de proposta i usuari siguin similars, i hi hagi interès, considerarem que la persona u està potencialment a favor de la proposta p . Quan els perfils de proposta i usuari siguin dispars, o no hi hagi interès, aleshores possiblement la proposta p no es recomanaria a la persona participant u :

$$A_u^p \sim \text{Bernoulli}((I_u^p + \sigma_2) \cdot (1 - \frac{|P^p - P_u|}{|P^p| + |P_u|})), \quad \sigma_2 \in (-1, 1)$$

- En cas de votació, el vot V_u^p de l'usuari a la proposta dependrà de l'alineació A_u^p i l'interès I_u^p de la persona usuària de la proposta. El valor que pugui prendre el vot serà binari, si vota a favor ($V_u^p = 1$) o no vota ($V_u^p = 0$), per tant, serà una variable que segueixi una Bernoulli. El paràmetre serà més gran (i, per tant, la probabilitat que voti) si hi ha interès i està a favor:

$$V_u^p \sim \text{Bernoulli}(\sigma_3 \cdot A_u^p \cdot I_u^p), \quad \sigma_3 \in [0, 1]$$

- Tant el número de seguidors S^p com el recompte R_u^p de vots de la proposta seran variables deterministes. Per una banda, S^p dependrà del nombre d'interessats i, per altra, R^p serà el nombre de vots a favor. Els resultats s'expressen:

$$S^p = \sum_{u \in \text{Usuaris}} \text{Bernoulli}(\theta \cdot I_u^p), \theta \in [0, 1]$$

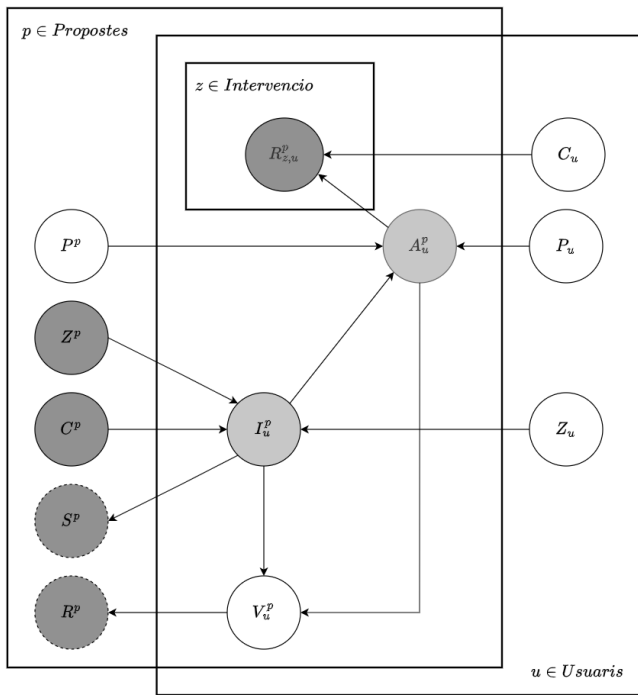
$$R^p = \sum_{u \in \text{Usuaris}} \text{Bernoulli}(\psi \cdot V_u^p), \psi \in [0, 1]$$

- Es considerava la variable aleatòria capacitat d'influència C_u de l'usuari segons la repercussió que tenien les seves intervencions. El valor que pugui prendre aquesta variable aleatòria serà binari, capacitat alta ($C_u = 1$) o baixa ($C_u = 0$), per tant, serà una variable que segueixi també una Bernoulli, el paràmetre de la qual serà p_1 :

$$C_u \sim \text{Bernoulli}(p_1)$$

- Donada una intervenció z de l'usuari u a la proposta p , la repercussió $R_{z,v}^p$ que tindrà dependrà de la capacitat d'influència C_u de l'usuari i l'alineació A_u^p entre usuari i proposta. La distribució que es proposa és una normal centrada al 0, de manera que sigui la corba de la normal d'on s'obtingui informació, essent més afilada com més capacitat d'influència s'observi o més polarització existeixi en l'alineació:

$$R_{z,u}^p \sim \mathcal{N}(0, 1/(|\sigma_4 - A_{z,v}^p| + C_u)), \sigma_4 \in [0, 1]$$



Per $p \in Propostes$:

$$P^p \sim \mathcal{N}(\mu_1, 1/\eta_1)$$

$$Z^p : P(Z^p = z_i) = \alpha_i, 1 \leq i \leq n-1, P(Z^p = z_n) = 1 - \sum_{i=1}^{n-1} \alpha_i$$

$$C^p : P(C^p = c_i) = \xi_i, 1 \leq i \leq m-1, P(C^p = c_m) = 1 - \sum_{i=1}^{m-1} \xi_i$$

$$S^p = \sum_{u \in Usuaris} \text{Bernoulli}(\theta \cdot I_u^p), \theta \in [0, 1]$$

$$R^p = \sum_{u \in Usuaris} \text{Bernoulli}(\psi \cdot V_u^p), \psi \in [0, 1]$$

Per $u \in Usuaris$:

$$P_u \sim \mathcal{N}(\mu_2, 1/\eta_2)$$

$$Z_u : P(Z_u = z_i) = \beta_i, 1 \leq i \leq n-1, P(Z_u = z_n) = 1 - \sum_{i=1}^{n-1} \beta_i$$

$$C_u \sim \text{Bernoulli}(p_1)$$

Per $u \in Usuaris$ i $p \in Propostes$:

$$I_u^p \sim \text{Bernoulli}(P(Z^p = z) * P(Z_u = z) + P(C^p = c) - \sigma_1), \sigma_1 \in [-2, 1]$$

$$A_u^p \sim \text{Bernoulli}((I_u^p + \sigma_2) \cdot (1 - \frac{|P^p - P_u|}{|P^p| + |P_u|})), \sigma_2 \in (-1, 1)$$

$$V_u^p \sim \text{Bernoulli}(\sigma_3 \cdot A_u^p \cdot I_u^p), \sigma_3 \in [0, 1]$$

Per $z \in Intervencio$:

$$R_{z,u}^p \sim \mathcal{N}(0, 1/(|\sigma_4 - A_{z,u}^p| + C_u)), \sigma_4 \in [0, 1]$$

Figura 4.7: Model gràfic probabilístic resultant de la plataforma decidim.barcelona

Capítol 5

Conclusions i futur treball

5.1 Conclusions

En aquest treball s'ha presentat un estudi emmarcat en la definició d'una comunitat virtual i la formalització de la figura d'un xatbot, proposant posteriorment un model basat en aprenentatge automàtic que permeti donar suport a les interaccions entre les persones participants de la comunitat virtual `decidim.barcelona`.

En una primera part, s'han identificat totes les variables en les quals es pot descompondre una comunitat virtual, proposant una estructura de com es relacionen entre elles. Posteriorment, s'introdueix la figura del xatbot, amb el que les persones usuàries poden entaular diàlegs.

En una segona part, el treball es basa en la teoria de models gràfics probabilístics per desenvolupar un model d'aprenentatge que permeti al xatbot resoldre consultes del tipus: “Quines persones usuàries tenen un perfil similar?”, “Quina proposta activa podria interessar a una persona participant”, “Quines persones usuàries tenen més capacitat d'influència?”, etc.

La hipòtesi d'aquest treball ha estat que seria possible construir un model que permeti a les persones usuàries de la comunitat virtual de `decidim.barcelona` entaular diàlegs amb un xatbot que li recomanés o informés de votacions actives de propostes que potencialment puguin resultar de l'interès de la persona usuària, posar en comú persones participants que potencialment tinguin perfils similars, etc. Mitjançant els recursos a l'abast de la plataforma, el model estimaria totes aquestes variables, garantint una estimació eficient d'aquestes.

L'estudi sistemàtic i la proposta d'una tècnica com la que es planteja en aquest estudi podria ajudar a impulsar la interacció en comunitats virtuals, més concretament a la comunitat virtual de `decidim.barcelona` amb la incorporació d'un xatbot.

5.1.1 Consecució d'objectius

L'objectiu general del treball plantejat consistia en formalitzar el concepte de comunitat virtual i especificar un xatbot dins la comunitat per tal de millorar l'experiència de la persona usuària, desenvolupant un model aplicable per generar més interacció. No obstant, això no impedeix apreciar que és un treball inacabat i que podria ser millorat aportant el comportament del model a mesura que es realitza inferència.

En quant als objectius parcials, més concrets, es podria afirmar que:

1. S'ha formalitzat el concepte de comunitat virtual, fent una revisió també a l'estat de l'art, i s'ha formalitzat la figura del xatbot.
2. S'han aportat exemples reals d'aplicació de la formalització, proposant interaccions amb xatbots amb l'objectiu d'aportar més interacció.
3. S'ha desenvolupat un model d'una part de la comunitat virtual `decidim.barcelona` basat en la teoria de models gràfics probabilístics, amb el qual es poden apropar més individualment propostes a la ciutadania.

5.2 Línies de futur treball

A la conclusió d'aquest treball, són diverses les línies d'investigació que queden obertes i podrien ser abordades en un futur.

- Fer inferència sobre el model, veient com es comporta amb la introducció de dades reals. Modificar variables, distribucions o paràmetres per aportar eficiència.
- La majoria dels xatbots canvien al llarg del temps quan s'afegeixen noves funcions o es requereixen nous serveis. Un repte pot ser dissenyar bots que canviïn amb el temps, no només en la seva aparença, característiques, personalitat, etc., sinó també en el rol en el què la comunitat els accepta. De la mateixa manera que un membre humà d'una comunitat pot guanyar més responsabilitats o més confiança al llarg del temps a mesura que aprèn i creix, els xatbots podrien esdevenir més o menys centrals per a la comunitat a mesura que s'hi adapten.
- Es podria definir tot un sistema d'agents que desenvolupessin la tasca de generar interacció, assignant-los a cadascun d'ells unes propietats que facilitessin la connexió amb els membres, uns rols que poguessin desenvolupar un o diferents agents i unes accions que permetrien la comunicació agent - usuari.

Bibliografia

- [1] [https://en.wikipedia.org/wiki/Convention_\(norm\)](https://en.wikipedia.org/wiki/Convention_(norm)).
- [2] [https://en.wikipedia.org/wiki/Ontology_\(information_science\)](https://en.wikipedia.org/wiki/Ontology_(information_science)).
- [3] <https://en.wikipedia.org/wiki/Role>.
- [4] Pablo Almajano. Assisted hybrid structured 3d virtual environments. 2014.
- [5] Theo Araujo. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85:183–189, 2018.
- [6] Jonathan P Bowen. Online communities: Visualization and formalization. *arXiv preprint arXiv:1307.6145*, 2013.
- [7] Gary Burnett. Information exchange in virtual communities: a typology. *Information research*, 5(4), 2000.
- [8] Kerstin Dautenhahn, Sarah Woods, Christina Kaouri, Michael L Walters, Kheng Lee Koay, and Iain Werry. What is a robot companion-friend, assistant or butler? In *2005 IEEE/RSJ international conference on intelligent robots and systems*, pages 1192–1197. IEEE, 2005.
- [9] Ajuntament de Barcelona. Decidim.barcelona. <https://www.decidim.barcelona/>.
- [10] Aldo De Moor and Hans Weigand. Formalizing the evolution of virtual communities. *Information Systems*, 32(2):223–247, 2007.
- [11] Aldo De Moor and Hans Weigand. Formalizing the evolution of virtual communities. *Information Systems*, 32(2):223–247, 2007.
- [12] Friederike Eyssel, Dieta Kuchenbrandt, Simon Bobinger, Laura De Ruitter, and Frank Hegel. 'if you sound like me, you must be more human' on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 125–126, 2012.
- [13] Gobierno de Aragón. <https://opendataei2a.aragon.es/servicios/chatbot/>, Febrer de 2022.

- [14] Ashok K Goel and Lalith Polepeddi. Jill watson: A virtual teaching assistant for online education. In *Learning engineering for online education*, pages 120–143. Routledge, 2018.
- [15] HN Io and CB Lee. Chatbots and conversational agents: A bibliometric analysis. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 215–219. IEEE, 2017.
- [16] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [17] Carlene Lebeuf, Alexey Zagalsky, Matthieu Foucault, and Margaret-Anne Storey. Defining and classifying software bots: a faceted taxonomy. In *2019 IEEE/ACM 1st international workshop on bots in software engineering (BotSE)*, pages 1–6. IEEE, 2019.
- [18] Hsiu-Fen Lin. The role of online and offline features in sustaining virtual communities: an empirical study. *Internet Research*, 2007.
- [19] Irene Lopatovska and Harriet Williams. Personification of the amazon alexa: Bff or a mindless companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 265–268, 2018.
- [20] Yukiko Nakano, Michael Neff, Ana Paiva, and Marilyn Walker. *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings*, volume 7502. Springer, 2012.
- [21] Timo Partala, Veikko Surakka, and Jussi Lahti. Affective effects of agent proximity in conversational systems. In *Proceedings of the third Nordic conference on Human-computer interaction*, pages 353–356, 2004.
- [22] Manuel Portela. Interfacing participation in citizen science projects with conversational agents. *Human Computation*, 8(2):33–53, 2021.
- [23] Joseph Seering, Michal Luria, Geoff Kaufman, and Jessica Hammer. Beyond dyadic interactions: Considering chatbots as community members. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.
- [24] Dhawal Shah. <https://www.class-central.com/report/mooc-stats-2016/>. *Class Central*, 2016.
- [25] Ella Tallyn, Hector Fried, Rory Gianni, Amy Isard, and Chris Speed. The ethnobot: Gathering ethnographies in the age of iot. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.
- [26] Barry Wellman and Milena Gulia. Virtual communities as communities. *Communities in cyberspace*, pages 167–194, 1999.

- [27] Barry Wellman, Janet Salaff, Dimitrina Dimitrova, Laura Garton, Milena Gullia, and Caroline Haythornthwaite. Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual review of sociology*, 22(1):213–238, 1996.
- [28] Etienne Wenger. *Communities of practice: A brief introduction*. 2011.

Apèndix A

Preliminars

A.1 Teoria de probabilitat

En aquest apartat es presentaran les definicions més introductòries, així com els principals resultats, de teoria de la probabilitat.

Definició A.1.1. Un espai de probabilitat és la terna (Ω, \mathcal{A}, P) on

- Ω és un conjunt que correspon al dels resultats de l'experiència aleatòria. S'anomena *espai mostral*
- \mathcal{A} és una família de parts d' Ω que té estructura d' σ -àlgebra, és a dir,
 1. $\Omega \in \mathcal{A}$
 2. Si $A \in \mathcal{A}$, també $A^C \in \mathcal{A}$, on A^C és el complementari de A , és a dir, $A \cup A^C = \Omega, A \cap A^C = \emptyset$
 3. Si $\{A_n, n \geq 1\} \subset \mathcal{A}$, es compleix $\bigcup_{n \geq 1} A_n \in \mathcal{A}$

La σ -àlgebra serveix per descriure tots els esdeveniments possibles relacionats amb l'experiència aleatòria.

- L'aplicació anomenada *probabilitat* determina l'assignació de versemblança dels esdeveniments:

$$P : \mathcal{A} \longrightarrow [0, 1]$$

i té les propietats següents:

1. $P(\Omega) = 1$
2. σ -*additivitat* Si $\{A_n, n \geq 1\}$ és una successió de conjunts de \mathcal{A} disjunts dos a dos, aleshores

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Com a conseqüència de la definició s'obtenen algunes propietats com:

1. $P(\emptyset) = 0$
2. Per a tot $A \in \mathcal{A}$, $P(A^C) = 1 - P(A)$
3. Si $A, B \in \mathcal{A}$, aleshores $P(A \cup B) + P(A \cap B) = P(A) + P(B)$

De vegades, es pot disposar d'informació complementària que permeti modificar el model probabilístic associat a una experiència aleatòria. Per tractar aquests tipus de qüestions s'introdueix el concepte de probabilitat condicionada.

Definició A.1.2. Sigui (Ω, \mathcal{A}, P) un espai de probabilitat i siguin $A, B \in \mathcal{A}$ dues experiències aleatòries, amb la probabilitat del conjunt B no nul·la. La probabilitat d' A condicionada per B es defineix per

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

D'aquesta definició obtenim directament $P(A \cap B) = P(A) \cdot P(B|A)$. Aquesta igualtat és coneguda com la regla de la cadena per probabilitats condicionades. Més generalment, si A_1, \dots, A_k són experiències aleatòries,

$$P(A_1 \cap \dots \cap A_k) = P(A_1) \cdot P(A_2|A_1) \cdot \dots \cdot P(A_k|A_1 \cap \dots \cap A_{k-1})$$

Definició A.1.3. Es diu que dos esdeveniments $A, B \in \mathcal{A}$ són independents si

$$P(A \cap B) = P(A)P(B)$$

Notació 3. Quan dos esdeveniments $A, B \in \mathcal{A}$ són independents ho notarem $(A \perp B)$

Sigui \mathcal{B} la σ -àlgebra de Borel, generada pels conjunts oberts de \mathbb{R} . Donat un espai de probabilitat (Ω, \mathcal{A}, P) ,

Definició A.1.4. Una variable aleatòria és una aplicació $X : \Omega \rightarrow \mathbb{R}$ que compleix

$$\forall B \in \mathcal{B}, X^{-1}(B) \in \mathcal{A}$$

Així, s'ha establert una assignació numèrica als elements de l'espai mostral o resultat de l'experiència aleatòria. També es pot transferir la probabilitat:

Definició A.1.5. La llei d'una variable aleatòria X és la probabilitat sobre $(\mathbb{R}, \mathcal{B})$ definida de la manera següent:

$$\forall B \in \mathcal{B}, P(X^{-1}(B)),$$

on $X^{-1}(B) = \{\omega, X(\omega) \in B\}$

La llei d'una variable aleatòria està fortament relacionada amb el concepte de funció de distribució:

Definició A.1.6. La funció de distribució associada a una variable aleatòria X és la funció $F : \mathbb{R} \rightarrow [0, 1]$ definida per

$$F(x) = P \circ X^{-1}((-\infty, x])$$

Distingirem dos tipus de variables aleatòries, les discretes i les contínues:

Definició A.1.7. Una variable aleatòria X direm que és discreta si la seva llei està concentrada en un conjunt numerable $B_0 \in \mathcal{B}$.

Com a exemple de variables aleatòries discretes destaca la Bernoulli: una experiència aleatòria amb dos resultats possibles a i b amb probabilitats respectives p i $1 - p$, $p \in [0, 1]$, és a dir, si tenim una variable X definida com una Bernoulli, llavors $P(X = a) = p$, $P(X = b) = 1 - p$ i ho indicarem $X \sim \text{Bernoulli}(p)$.

Definició A.1.8. Una funció $f : \mathbb{R} \rightarrow \mathbb{R}$ s'anomena una densitat si compleix les condicions següents:

1. $f \geq 0$,
2. f és integrable (en el sentit de Riemann) en \mathbb{R} ,
3. es té que

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Definició A.1.9. Es diu que una variable aleatòria X és absolutament contínua (o té llei absolutament contínua) amb densitat f si la seva funció de distribució F es pot escriure com

$$F(x) = \int_{-\infty}^x f(y) dy,$$

per tot $X \in \mathbb{R}$, on la funció f satisfà les condicions de densitat.

Com a exemple de variable aleatòria absolutament contínua destaca la distribució normal. La funció de densitat d'una variable aleatòria X és:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

Aquesta funció depèn de dos paràmetres, la mitjana μ , i la variància σ .

La independència de les variables aleatòries X_1, \dots, X_n voldrà dir que els valors presos per una variable qualsevol no afecta els de les altres.

Definició A.1.10. Siguen X_1, \dots, X_n variables aleatòries definides en un espai de probabilitat (Ω, \mathcal{A}, P) . Direm que són independents si per a conjunts qualssevol $B_1, \dots, B_n \in \mathcal{B}$, es compleix que

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i)$$

En el cas de variables aleatòries discretes la condició d'independència s'escriu de la manera equivalent següent:

Per a qualsevol $x_1, \dots, x_n \in \mathbb{R}$,

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$$

Observació A.1.11. La propietat d'independència és pot aplicar sobre probabilitat condicionada. Siguin X, Y, Z variables aleatòries de (Ω, \mathcal{A}, P) . Direm que X i Y seran independents donat Z si per a conjunts qualsevol $B_1, B_2, B_3 \in \mathcal{B}$ es compleix que

$$P(X \in B_1, Y \in B_2 \mid Z \in B_3) = P(X \in B_1 \mid Z \in B_3) \cdot P(Y \in B_2 \mid Z \in B_3)$$

i ho notarem $(X \perp Y \mid Z)$

A.2 Teoria de grafs

En aquesta secció, analitzem alguns dels conceptes bàsics de la teoria de grafs utilitzats al treball:

Definició A.2.1. Un graf $\mathcal{K} = (V, E)$ està format per un conjunt d'elements V , anomenats vèrtexs i un conjunt d'arestes E , que uneixen dos vèrtexs. Dos vèrtexs $X_i, X_j \in V$ poden estar connectats per una arista dirigida $X_i \rightarrow X_j$, o per una no dirigida $X_i - X_j$. Si un graf té totes les arestes dirigides, l'anomenarem graf dirigit.

Definició A.2.2. Direm que $X_1, \dots, X_k \in V$ formen un camí dirigit en el graf $\mathcal{K} = (V, E)$ si, per cada $i = 1, \dots, k - 1$ tenim que $X_i \rightarrow X_{i+1} \in E$.

Definició A.2.3. Direm que $X_1, \dots, X_k \in V$ formen una pista en el graf dirigit $\mathcal{K} = (V, E)$ si, per cada $i = 1, \dots, k - 1$ tenim que $X_i \rightarrow X_{i+1} \in E$, o bé, $X_{i+1} \rightarrow X_i \in E$

Definició A.2.4. Un graf estarà connectat si per cada $X_i, X_j \in V$ existeix una pista entre X_i i X_j .

Definició A.2.5. Direm que $X \in V$ és un ancestre $Y \in V$ a $\mathcal{K} = (V, E)$, i que Y és un descendents X , si existeix un camí dirigit X_1, \dots, X_k amb $X_1 = X$ i $X_k = Y$. S'utilitza $Desc_X$ per denotar els descendents d' X , Anc_X per denotar els ancestres d' X i $NoDesc_X$ per denotar el conjunt $\mathcal{X} \setminus Desc_X$

Definició A.2.6. Sigui $\mathcal{K} = (V, E)$ un graf. Un ordre dels nodes $X_1, \dots, X_n \in V$ és un ordre topològic relatiu a \mathcal{K} si, sempre que $X_i \rightarrow X_j \in E$, aleshores $i < j$.

Definició A.2.7. Un cicle en un graf $\mathcal{K} = (V, E)$ és un camí dirigit X_1, \dots, X_k on $X_1 = X_k$. Un graf és acíclic si no conté cicles.

Proposició A.2.8. Si $\mathcal{K} = (V, E)$ és graf dirigit acíclic, aleshores existeix, com a mínim, un ordre topològic.