

The Semantic Evolution of Online Communities

Matthew Rowe
School of Computing and Communications
Infolab21, Lancaster University,
Lancaster, UK
m.rowe@lancaster.ac.uk

Markus Strohmaier
University of Koblenz & GESIS
Untersachsenhausen 6-8
50667 Cologne
markus.strohmaier@gesis.org

ABSTRACT

Despite their semantic-rich nature, online communities have, to date, largely been analysed through examining longitudinal changes in social networks, community uptake, or simple term-usage and language adoption. As a result, the evolution of communities on a semantic level, i.e. how concepts emerge, and how these concepts relate to previously discussed concepts, has largely been ignored. In this paper we present a graph-based exploration of the semantic evolution of online communities, thereby capturing dynamics of online communities on a conceptual level. We first examine how semantic graphs (concept graphs and entity graphs) of communities evolve, and then characterise such evolution using logistic population growth models. We demonstrate the value of such models by analysing how sample communities evolve and use our results to predict churn rates in community forums.

1. INTRODUCTION

A common characteristic of online communities is their capacity to create, discuss and evaluate information. To date, understanding the evolution of online communities has largely focussed on the examination of social network properties [3, 6, 10] and social-group formation [1], or examining how term-usage within communities changes longitudinally [4]. Such works afford insights into how the social structure of communities evolve and the terminology that develops from *neologisms* to standard language, however existing work has not considered how concepts within communities emerge, how the structure of concepts evolves over time, and how communities differ from one another with respect to both the emergence and evolution of semantics. In examining the *semantic evolution* of online communities one could understand how semantic development impacts user retention in communities, and arm community managers with key signals of their community becoming *unhealthy*.

In this paper we examine how online communities evolve, not socially, but *semantically* in terms of the concepts found

within discussion threads, using community data from the Irish community message board Boards.ie.¹ We present an approach that models time-delimited semantics within online communities as semantic graphs, containing interlinked concepts (DBPedia Ontology classes) and entities (DBPedia Resource URIs). Such an approach allows one to inspect the relations between the entities and concepts discussed by the community, and the density of semantic structures, thereby informing how semantics appear within the community (through *relations*) and how concepts fit in with prior concepts in the community (through *density*). Our contributions are three-fold, we present: (i) a graph-based approach to characterise time-dependent community semantics; (ii) an approach to model semantic graph evolution using logistic population models (capturing growth rates and capacities); and (iii) applications of logistic model dynamics to analyse the semantic evolution of communities, and predict the churn rate from the communities.

Our findings indicate that one can predict social characteristics of communities, e.g. *churn rate*, based on communities' semantic evolution. This work thereby lays a basis for deepening our understanding of how social and semantic layers interact with each other in the evolution of online communities.

2. RELATED WORK

One of the most fundamental works to examine how online communities change and evolve was done by Iriberry and Leroy [7] in which the authors proposed, after reviewing several pieces of work in the realm of online community management, online community *lifecycles* that involve five stages: *inception, creation, growth, maturity, and death*. The work of Belak et al. [2] examined how *offline* scientific communities changed and evolved, focussing on the merging and birth of new communities - analogous to the work of Iriberry and Leroy [7] in defining virtual community lifecycles. Danescu et al. [4] assessed the changing terminology of users with respect to communities over time, examining how users tend to adapt the language of the community at first before trailing off to not use this. The authors inspected the terms in the titles of publications and how they relate to users joining communities - conferences in the case of DBLP.

Changes in social networks and their dynamics have been examined in a number of works: Mislove et al. compared the structural properties of Flickr, LiveJournal, Orkut and YouTube [10] by examining various characteristics (link symmetry, power law distributions of edges and nodes, and local

¹<http://www.boards.ie>

clustering of users) and found high degrees of local clustering on the different platforms which contained densely populated subgroups of similar users. Recent work by Gong et al. [6] inspected the evolution of social networks on Google+ as the platform was growing in memberships, in particular they focused on *social-attribute networks* (i.e. bipartite graphs containing people and their attributes as nodes), finding that the platform exhibited unique growth and characteristics of the networks as more people joined Google+. Leskovec et al. [8] modelled the development of social networks across four platforms (Flickr, Delicious, Yahoo! Answers and LinkedIn) by modelling the process of node arrival (users joining), edges being created and waiting times between edge creation. Previous work mostly ignored term based and semantic information and concentrated on how the social networks evolved, not the communities. Although such works afford insights into the evolution of social networks, they do not consider how a community of users evolve semantically.

3. CHARACTERISING ONLINE COMMUNITIES WITH SEMANTIC GRAPHS

For our experiments we used data from the Irish community message board Boards.ie.² This is a general-discussion community message board that includes a set of hierarchically nested forums (F) in which posts are made - i.e. forum A can be a parent of forum B, and thus B contains specialised topics of discussion over A. Posts are provided as a set of quadruples $\langle u, s, t, f \rangle \in P$, where user u posted message s at time t in forum f . A message (s) is composed of terms that we use to build the semantic models for individual communities. The information discussed in a community, and thus its semantics, can change and alter over time, therefore we constrain a community's model to specific time snapshots - e.g. $t' \rightarrow t''$ where $t' < t''$ - for this we use the following construct that filters through all relevant posts' contents within the allotted time window:

$$S^{t't''} = \{s : \langle u, s, t, f \rangle \in P, t' \leq t < t''\} \quad (1)$$

Information discussed within online communities can be represented in terms of its semantics, using information from either the schema-level (i.e. ontological classes and relations between them) or the data level (i.e. using entities and how they are related to one another). For the former we consider *concepts* to be classes found within the DBpedia Ontology, that is: the types of entities that users are discussing (e.g. people, locations, etc.), while for the latter case we consider DBpedia resources: i.e. entities themselves (e.g. `dbpedia:Barrack_Obama`). Given our set of post contents, $S^{t't''}$, we derive concepts and entities from a forum over a time period as follows: we process each post content $s \in S^{t't''}$ using an entity extraction tool $\Psi(s)$ to return the set of entities related to the content of s . Given the entities (R_E) returned for a given community forum over an allotted time period we then construct two types of *semantic resource graphs*: *concept graphs*, which function at the schema-level and contain class information; and *entity graphs*, which function at the data level and contain information that relates entities to one another.

²<http://www.boards.ie>

3.1 Concept Graphs

A concept graph (G_C) is a type of semantic resource graph that contains the types of entities found within a given forum as vertices (V) and the relations between these classes as edges (E). For a given community forum f we have a set of entities R_E that were extracted over some period time $t' \rightarrow t''$. Each entity, given that we are using DBpedia resource URIs, is typed according to one or more classes from the DBpedia ontology.³ Therefore to construct the set of concepts that are cited within a given forum f over the allotted time period we retrieve the classes that each entity is a type of and store these in the following set: R_C . From this set we generate a time-dependent concept graph: $G_C[f, t', t''] = \langle V_f, E_f \rangle$, such that $G_C[f, t', t''] \subset G_{type}$ - where G_{type} denotes the DBpedia type graph formed from the class structure of the DBpedia ontology. In this context the set of concepts denotes the seed set and is used to populate the vertices in the graph and then construct edges between the vertices based on existing links between the concepts in the DBpedia type graph (G_{type}):

$$E_f = \{(c_i, c_j) : c_i, c_j \in V_f, (c_i, c_j) \in E_{type}\} \quad (2)$$

In order to derive the set of vertices we must consider how the seed set can be used for this process as it is often the case that the set is comprised of concepts which are not directly connected to one another in the concept graph. To connect such concepts, and derive a fully connected concept graph (i.e. with no disconnected components) we extract the *Root Path Graph* as follows: From each concept ($c \in R_C$) we identify the parent concept (`<c rdfs:subClassOf p>`) and iteratively move up the concept graph until the root node is reached (`owl:Thing`), thereby returning a set of nodes that formed the path from c to the root node: $rootpath(c) = \{c, p, \dots, owl:Thing\}$. The graph's vertices are therefore derived by taking the union of all paths to the root returned from each concept within the seed set:

$$V_f = \bigcup_{c \in R_C} rootpath(c) \quad (3)$$

3.2 Entity Graphs

An entity graph (G_E) is a type of semantic resource graph where the vertices (V) are entities and the set of edges (E) connecting these entities are relations between them derived from the Web of Linked Data. We define the entity graph as $G_E[f, t', t''] = \langle V_f, E_f \rangle$, such that $G_E[f, t', t''] \subset G_{entity}$ - where G_{entity} denotes the DBpedia entity graph containing relations between entities at the data level. As we are provided with a collection of time-delimited entities R_E for a given forum, we query DBpedia for links between entity pairs and add such edges to the graph where such a link exists:

$$E_f = \{(r_i, r_j) : r_i, r_j \in V_f, (r_i, r_j) \in E_{entity}\} \quad (4)$$

Given this edge construction mechanism we only look for relations one-hop away in the entity graph, that is: given R_E we only look for relations between elements in the set. This could be extended to include 2-hop relations, however we are interested in how entities in the communities are connected to one another *directly*. In this work, the vertices in the entity graph are thus those entities which are found to be connected to one another directly via 1-hop distances.

³<http://dbpedia.org/Ontology>

3.3 Graph Measures

Having defined time-delimited concept graphs and entity graphs, we can then inspect dynamics of the graph to see how it is evolving over time, and thus how information within a given community is developing. To perform such inspections we use five graph-measures from network analysis:

1. *Node Count*. This measure is intended to capture the *size* of the graph.
2. *Diameter*. The longest shortest path between any pair of nodes, in the graph; capturing the *breadth* of the graph.
3. *Specialisation Count*. Specific to concept graphs, this counts how many topic specialisations occurred in the community.
4. *Graph Entropy*. Measures the *density* of the graph based on the number of connections that flow from nodes:

$$H(G) = - \sum_{c \in V_f} p(c) \times \log p(c) \quad (5)$$

Where $p(c) = \text{outdeg}(c)/|E|$. This measure is inspired by work by Navigli and Lapata [11] in which the authors measure the entropy of word sense disambiguation subgraphs. In the context of our work, a higher entropy denotes greater graph density.

5. *Clustering coefficient*. Measures the *cliquishness* of the graph. A higher value indicates that a node's neighbours have a tendency to be well-connected. Let $\text{inout}(v_i)$ be the in-degree plus the out-degree of the node and a_{jk} denote an entry in the graph's adjacency matrix (1 for an edge connecting node j and k and 0 otherwise), then the clustering coefficient is:

$$CLC(v_i) = \frac{\sum_{j \in V_i^1} \sum_{k \in V_i^1} a_{jk}}{\text{inout}(v_i)} \quad (6)$$

We derive the global clustering coefficient for a graph from the average of local clustering coefficients for all nodes.

The concept graphs constructed in this work are *fully connected*; we opted for a fully-connected graph given the relatively small number of potential nodes (classes) that could belong in a single component. The entity graph, however, could potentially contain several disconnected components as we only consider edges between the derived entities that appear within the DBPedia entity graph; hence, the above measures are calculated for the entity graph over the *largest* connected component for a given community forum.

4. EXAMINING MACRO EVOLUTION

We first examine how communities on Boards.ie evolve on a macro level, in order to understand what processes and behaviour are occurring, before using such insights to choose a suitable model and fit this to individual community forums. To perform our experiments we used data from Boards.ie following 23-04-2005 until the end of the dataset (08-02-2008). For each forum that was active during that period (> 1 post per week giving 93 forums to analyse) we segmented the forum posts into one-week windows and then extracted entities from posts using the TextRazor entity extractor.⁴ Our

⁴<http://www.textrazor.com/>

decision to use TextRazor was based on: (i) prior work by Derczynski et al. [5], in which the authors found TextRazor to yield good performance, and; (ii) on the issue of scale, as we are handling three years' worth of data we needed an extractor that did not have API quotas. Following this process we had entities extracted from all posts (4.5m) covering a 150-week period. We segmented the first 120 weeks into an *analysis* period over which we examined the semantic evolution of the community forums, and used the remaining 30 weeks for a churn rate prediction experiment that we will discuss in a later part of the paper.

We examined the concept graphs and entity graphs of the online communities by forming the graphs from *cumulative segments* by keeping the start of the window at the same position (i.e. at $t = 0$) but extending the width of the window such that it is increased in size (and thus covers more weeks): $t_0 \rightarrow t_n$, where $n = \{1, 2, \dots, 120\}$. To examine the macro evolution of forums, we derived the graph measures for each community and then took the average of these measures.

4.1 Macro Evolution of Concept Graphs

Fig. 1 shows the macro evolution of the community concept graphs over time based on: node count, graph entropy and specialisation count, together with the 95% confidence intervals of the measures' means. We omitted diameter as this did not change drastically. By examining the cumulative graph we can inspect how it grew over time in terms of concepts arriving in the communities: providing some indication of the extent to which new concepts arise. As seen in Fig. 1(a), the classes of entities cited in the communities grows in number quickly before the rate of growth tapers off and appears to converge to a given limit. This suggests that although a constant number of concepts are discussed within community forums over time, they are not entirely distinct such that after time the arrival of new, novel concepts is somewhat limited and thus the growth rate slows. When we examine the cumulative development of the concept graphs, the average *graph entropy* of concept graphs begins to converge on a limit (Fig. 1(b)) suggesting that the evolution of the concept graphs, as more nodes arrive cumulatively but at a reduced pace, grows in density but at a reduced pace as time passes. Similar to the node count and graph entropy, in Fig. 1(c) we see how the initial growth in *specialisations* is at a heightened rate before tapering off as the total number of specialisations converges to a limit. This suggests that conversations in the Boards.ie communities tend to specialise over time.

4.2 Macro Evolution of Entity Graphs

Inspection of the *node counts* in the entity graph indicates that the number of entities cited in the communities grows, on average, over time. The cumulative plot (Fig. 2(a)) indicates that the number of new nodes arriving in the graph cumulatively is relatively linear and fitting a linear regression model yields a maximum likelihood estimate for the slope of $\hat{\beta} = 9.332$ - thus, on average, 9 new entities are discussed in a community every week. In terms of the *diameter* of the entity graph, the graph increases over time and appears to be converging on a limit (Fig. 2(b)). Like the concept graph, we find that the *entropy* of the entity graph tends to an upper limit that it converges on. The *clustering coefficient* also tends to an upper limit over time, suggest-

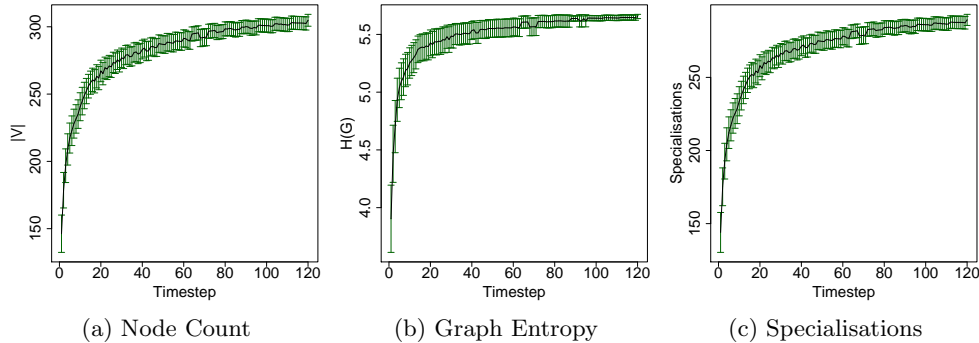


Figure 1: Concept graphs’ evolution based on node counts, graph entropy and specialisations.

ing that the density of the graph grows as more entities are added and thus more connections are possible between them.

Summary: We can summarise the following salient findings: (i) for concept graphs: node count, specialisation count and density (graph entropy) tend to converge to limit; (ii) for entity graphs, the diameter, graph entropy and clustering coefficient tend to converge to a limit, while the node count (number of entities) increases linearly; (iii) despite new entities arriving at a constant linear rate, on average, the number of concepts tends to converge on a maxima.

5. MODELLING SEMANTIC EVOLUTION

In the previous section, we found that the concept graphs and entity graphs tend to evolve in a convergent manner: that is, for different measures they tend to evolve towards a limit. Such limiting evolution has been found previously in population models where a given population has a *carrying capacity* that the population evolves towards at differing *proportionate growth rates*. These proportionate growth rates slow down over time as the population tends towards a limit (the carrying capacity) - we see this in tapering curves in the aforementioned graphs. An immediate question that arises from this effect is: *how do the communities differ in terms of evolution rates?* To answer this question we used logistic population models that contain: (i) the growth rate of the graphs (r), and (ii) the carrying capacity of the graphs (E). Each of these variables can be used to characterise the community forums (93 forums in total) in terms of their semantic evolution given a measure (e.g. node count in the concept graph). To derive the variables r and E for a given community forum and graph measure (m) we derive a set of time steps (T) which depict a change in a graph measure:

$$T = \{a : a \in [1, 119], m(G_{1,a+1}) > m(G_{1,a})\} \quad (7)$$

Deriving the set of change time steps for a given community allows the proportionate growth rate for a given time step ($t \in T$) to be derived: $R_t = (P_{t+1} - P_t)/P_t$. This value is equivalent to the following equation which defines the proportionate growth rate R_t in terms of the community’s growth rate (r) and carrying capacity (E), our unknown variables: $R_t = r(1 - P_t/E)$. Therefore if we measure the proportionate growth rate over the $|T|$ distinct time steps then we can derive, via simultaneous equations, the growth rate of the graph and its carrying capacity, the very measures that we can use to characterise the semantic evolution

of a given online community based on a single graph measure. We exclude the derivation of the equations from the paper, but it is sufficient to conclude that given $|T|$ time steps we would have a single equation for each time step ($t \in T$): $R_t r^{-1} + P_t E^{-1} = 1$. We can then solve for the unknown variables r and E using the QR-decomposition of a matrix: expressing the lefthand side of the simultaneous equations as a $|T| \times 2$ matrix and the righthand side as a $|T|$ -element vector where each element is 1. We induced logistic population models for each of the graph measures (aside from entity graph node count) and examined how the growth rate and carrying capacities were distributed - we found all models to be suitable fits at the 1% significance level using the chi-squared goodness-of-fit test.

We omitted the plots showing the distribution of *growth rates* and *carrying capacities* for communities’ concept and entity graphs’ measures, however these distributions demonstrated the following:

- All communities evolve to cover roughly 80% of the ontological classes (total number is 359), however some communities converge on the maximum quicker than others - demonstrated by a high growth rate for a small number of communities.
- The majority of communities’ concept graphs become denser at a slow rate, while a few communities’ become denser quickly, thus suggesting that users tend to discuss concepts that are orthogonal and not related to one another in the majority of communities.
- Communities’ entity graphs show variance in the rate by which entities are discussed for the first time within the communities, however such rates are linear: fitting a per-community linear regression model regressing the week count on the node count yielded a lowest coefficient of determination of 0.3 for one community.

6. APPLICATIONS

We now demonstrate the utility of graph-based approaches via two applications: community analysis, and churn rate prediction.

6.1 Community Analysis

To more closely examine the differences between community forums in terms of their semantic evolution we characterised each community in our dataset by its *semantic dynamics motif* over the 120-week analysis period: that is, the observed evolution measures derived from the concept

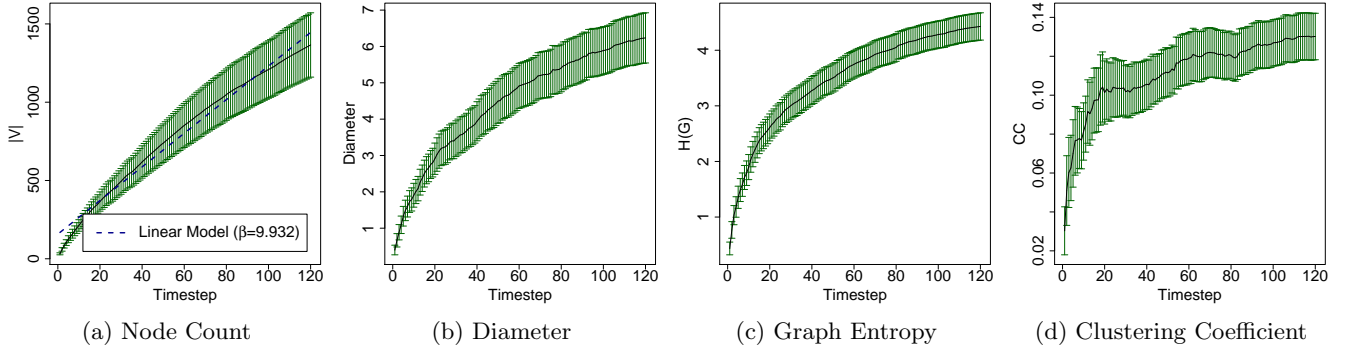


Figure 2: Entity graphs' evolution based on node counts, diameter, graph entropy, and clustering coefficient.

and entity graphs defined within a single vector for a given community (f): $\mathbf{m}_f = \{m_1, m_2, \dots, m_{13}\}$. We began by deriving a single matrix $\mathbf{M} \in \mathbb{R}^{93 \times 13}$ containing the 93 communities (rows) under analysis together with their 13 evolution measures (columns), and performing principal component analysis over this matrix. The result of this clustering is shown in Fig. 3 where we have colour-coded the different community forums by their hierarchical level in the platform (level 2 = most general, level 4 = most specific). We found that several of the more general forums appeared as outliers in the plot and thus exhibited unique evolution dynamics, while the two level 4 forums (forum 556 and 554) were bunched together suggesting that they follow similar trends. We further examined the semantic motifs of three outlier communities from differing levels:

- *Level 2: Forum 7 - After Hours.* A random discussion forum.
- *Level 3: Forum 227 - Television.* Discussions about television.
- *Level 4: Forum 554 - Wanted Motors.* Discussions about cars and car parts.

Fig. 3 presents clear differences between the forums: we note that for the concept graph dynamics (CG) the rate of the node count is lowest for the After Hours forum but that the node count equilibrium is highest, indicating that the more general the forum the slower is the growth of the concept graph, but the higher the maxima of the graph size. Likewise for the specialisation count in the concept graphs: the After Hours forum exhibits a slower rate of growth but with a greater carrying capacity that the graph is tending towards. In terms of the entity graph: the slope of node count growth described by the linear model is highest for After Hours, indicating that compared to the other two forums, the rate at which new entities are cited by the community of users is much greater, while for the more topically-specific forum of the Wanted Motors forum this is a lot lower. The entity graph equilibrium is also highest for After Hours and lowest for Wanted Motors, indicating that the more general a forum is the greater the carrying capacity of its entity graph and the greater the number of entities that will be discussed.

6.2 Churn Rate Prediction

To examine the link between the semantic evolution of online communities and their social properties, we defined a prediction task in which we used the semantic evolution

dynamics of a given community at time step t to predict the churn rate of community members at time $t + 1$. We defined the *churn rate* of a community as the proportion of active users during a given time period (i.e. week segment) that post for the last time. We used the *semantic dynamics motifs* from the prior experiment (as listed within Fig. 3) and also included graph measures at a given time period: i.e. graph entropy at time t , specialisation count at time t , etc. We derived these features for every time step for each community and derived the response variable as the churn rate at the following time step. We then compiled a training dataset (up to week 120) and a test dataset (from week 120). Each dataset had the following form: $D = \{(\mathbf{x}_i, y_i)\}$, where \mathbf{x}_i contained a 21-element time-delimited feature vector for a given community and y_i was the churn rate of the community at the following time step. We trained a ridge regression model (ψ) using D_{train} and applied it to D_{test} , testing the performance of: a) just concept graph features, b) just entity graph features, and c) all features. An autoregressive model was used as the baseline - using the churn rate at time t as a single predictor variable for the churn rate at time $t + 1$. Performance was evaluated using the Root Mean Square Error (RMSE).

Table 1: Root Mean Square Error when predicting churn rates using: an Autoregressive model ($R^2 = 0.341$) and a Ridge Regression model using Concept Graph, Entity Graph, and all features.

Baseline	Concept Graph	Entity Graph	All Features
7.310×10^{-3}	5.315×10^{-3}	5.301×10^{-3}	4.941×10^{-3}

Table 1 presents the results from our prediction our experiment. We found that for all tested models (concept graph, entity graph, all features) we significantly outperformed the baseline - tested using the sign test ($\alpha = 0.001$). Entity graph features outperform concept graphs but not significantly, while our best model is the use of all features together in a single model. These results empirically demonstrate the utility of semantic evolution dynamics in predicting community churn rates, and suggests a link between how the communities develop semantically and the likelihood of users leaving the communities.

7. DISCUSSION & CONCLUSIONS

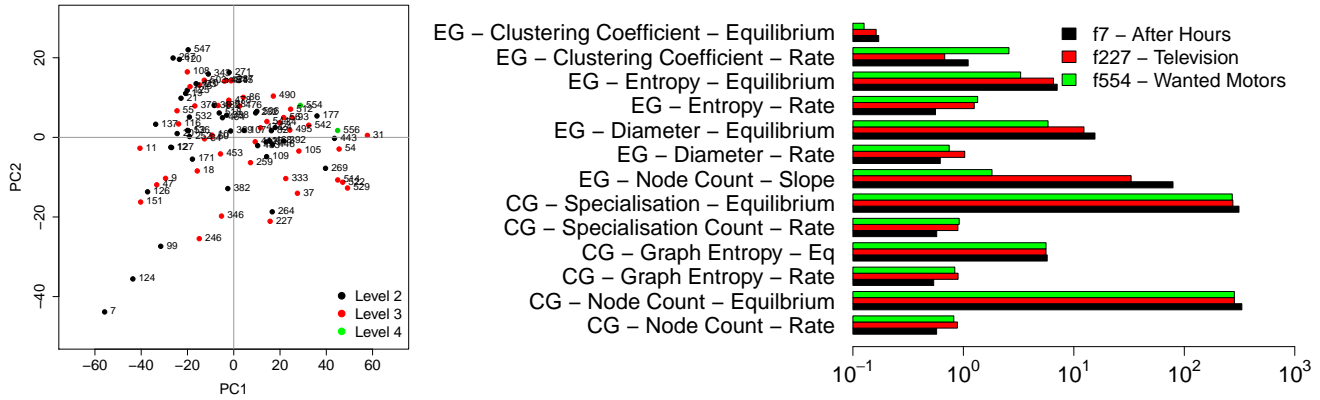


Figure 3: PCA plot of the communities based on their semantic motifs (left) where level 4 forums are clustered together, and model values for the concept graphs (CG) and entity graphs (EG) for the three outlier forums from the three levels (right).

In this work, we found that concept and entity graph density in boards.ie does not grow linearly (unlike in social networks [8]) but instead converges on a limit, which we characterised as the carrying capacity (E) of a given community’s concept and entity graph entropy. We also discovered that the diameter of the entity graph in our online community converged on a limit over time as the rate of concepts arrived slowed down, again in contrast to findings from the social networking domain where diameters were found to shrink as more nodes joined the network [9]. Indeed, this notion of convergence to a limit is common across all but one of the graph measures that we examined and suggests that online communities have a finite number of topics that can be discussed and that semantics will converge on a maxima over time.

Our contributions can be summarized as follows: (i) We used semantic graphs to firstly examine how concepts discussed by communities changed over time at a macro-level, (ii) we used logistic population models to inspect how individual communities evolved over time, and (iii) we deployed logistic population models to capture semantic graph changes along different measures and applied our results to community analysis and churn rate prediction. Thereby, our work forms a basis for combining studies of social and semantic network evolution in future work.

8. REFERENCES

- [1] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54. ACM, 2006.
- [2] Václav Belak, Marcel Karnstedt, and Conor Hayes. Life-cycles and mutual effects of scientific communities. *Procedia - Social and Behavioral Sciences*, 22(0):37 – 48, 2011.
- [3] Kon Shing Kenneth Chung, Mahendra Piraveenan, and Shahadat Uddin. Community evolution and engagement through assortative mixing in online social networks. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 0:724–725, 2012.
- [4] Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the World Wide Web Conference*, 2013.
- [5] Leon Derczynski, Diana Maynard, Niraj Aswani, and Kalina Bontcheva. Microblog-genre noise and impact on semantic annotation accuracy. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media (HT 2013)*, 2013.
- [6] Neil Zhenqiang Gong, Wenchang Xu, Ling Huang, Prateek Mittal, Emil Stefanov, Vyas Sekar, and Dawn Song. Evolution of social-attribute networks: Measurements, modeling, and implications using google+. *CoRR*, abs/1209.0835, 2012.
- [7] Alicia Iriberry and Gondy Leroy. A life-cycle perspective on online community success. *ACM Comput. Surv.*, 41(2):11:1–11:29, February 2009.
- [8] Jure Leskovec, Lars Backstrom, Ravi Kumar, and Andrew Tomkins. Microscopic evolution of social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 462–470. ACM, 2008.
- [9] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, KDD ’05, pages 177–187, New York, NY, USA, 2005. ACM.
- [10] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *SIGCOMM conference on Internet measurement*, IMC ’07, pages 29–42, 2007.
- [11] Roberto Navigli and Mirella Lapata. An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(4):678–692, 2010.