Li Lei, Limor Raviv and Phillip Alday
Radboud University Nijmegen and Max Planck Institute for Psycholinguistics, Nijmegen

RUSE 2019

Language Change in Online Social Networks

Over the past few decades, computational methods have emerged as a powerful tool in studying language change in social networks, given their ability to test hypotheses and easily manipulate social properties in large scale.

For example, a previous study (Fagyal, Swarup, Escobar, Gasser, & Lakkaraju, 2010) has proposed a degree-biased voter model (DBVM) to simulate language change in social networks, focusing on the role of hubs and loners in the community. There were eight linguistic variants which could spread in a bi-directional closed network and each agent in the network would choose a neighbour to update its linguistic variant. The selection rule was that a higher in-degree leads to a higher chosen probability. In the DBVM, the establishment of novel variants could be observed after language diffusion. This model revealed that loners play a key role in language change as variant-keepers.

However, the social network structure used in Fagyal et al. (2010) was simulated and based on "an artificial but socially realistic influence network", which relies on previous theories on social science and does not necessarily adhere to all properties of real-world networks. Here, we extend the original model to use network structure inspired by real-world online communities, and explore the model's application to such a network.

First, we successfully replicated the DBVM model presented in Fagyal et al. (2010), which included 900 agents and 7561 edges. We obtained a similar in-degree distribution (Figure 1a) and similar diffusion results over 40,000 iterations (Figure 1b), following the same pattern identified in Fagyal et al. (2010). The result showed that competing variants which were represented by different colours in Figure 1b could establish the norm alternately over time on the basis of different chosen probability.

Next, we tested this model using a real-world "who-trust-whom" online social network with 75,879 agents and 508,837 edges, which was extracted from a general review website Epinions (Richardson, Agrawal & Domingos, 2003). While its in-degree distribution was different from that of the original network (Figure 1c), the process of language change indicated a highly similar diffusion dynamics between two networks (Figure 1d). This result shows that the DBVM can model a real-world network.

Finally, we scaled down the Epinions network to include 900 agents with the same in-degree distribution and ratio and ran several simulations with different numbers of edges (5000, 7561, 27000). The results showed similar trends in language change: alternate norms of linguistic variants could be observed. Interestingly, we found that more edges lead to a longer and smoother fixation of linguistic variants. This result suggests that language change is more difficult in a high-density network, in line with an empirical study in Belfast's Protestant enclaves (Milroy, 1987): a norm is easier to maintain in a high-density community. In ongoing work, we plan to change additional variables of the online social network, such as a more "socially realistic" update rule instead of simple in-degree-bisad communication.

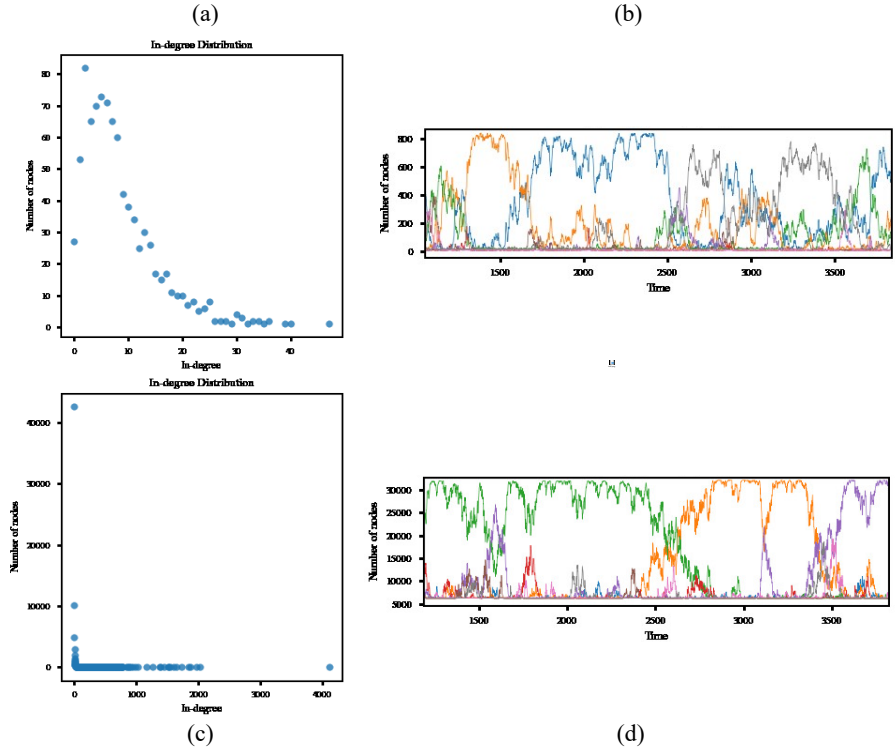

Figure 1: (a) In-degree distribution in the replicated DBVM; (b) Changes in the prominence of linguistic variants over an extended time series in the replicated DBVM; (c) in-degree distribution in the real-world Epinions network; (d) Changes in the prominence of linguistic variants over an extended time series in the real-world Epinions network.

References

Fagyal, Z., Swarup, S., Escobar, A. M., Gasser, L., & Lakkaraju, K. (2010). Centers and peripheries: Network roles in language change. *Lingua, 120*(8), 2061-2079.

Milroy, L. (1987). *Language and social networks*. Oxford: Basil Blackwell.

Richardson, M., Agrawal, R., & Domingos, P. (2003). Trust management for the semantic web. *Proceedings of the 2nd International Semantic Web Conference*. 351–368.