Dynamic network generative model

Habiba, Chayant Tantipathananandh, Tanya Berger-Wolf
University of Illinois at Chicago.

In this work we present a statistical model for generating realistic dynamic networks over time. Modeling such networks is necessary both to better understand the underlying dynamics of the population, as well as, for generating synthetic networks that emulate the real–world properties, for validating analysis.

Classically there are two main schools of network modeling. The approach primarily used in social sciences is to treat the entire network as a maximum likelihood object from a statistical distribution where some of the parameters, such as the number of dyads or triads, are fixed [4]. A radically different approach originates from the random graph community, where generative models are designed to emulate large scale global statistical behaviors in networks, such as the degree distribution [1, 2] and average distance [5], among others. A significant shortcoming of these classes of models is that they are fundamentally evolutionary but not truly dynamic. That is, once a connection is established it cannot be removed. This intuitively contradicts how social links are formed, reinforced, and lost in real world. These models essentially give a "static" representation of the dynamics of interactions aggregated up to a certain point in time.

Here is a simple example of what the above mentioned models fail to capture. At given point in time, society exists as a collection of loosely formed communities [3, 7]. As an individual joins a community, she/he forms relationships with its members. Overtime the individual updates those relationships by adding more links to already existing members or new–comers and by removing other links. Moreover, some individuals leave the communities all–together. Most of the models in the aforementioned two classes capture one or the other network properties but fail to incorporate many others. Especially, the existing models do not capture dynamic of forming and breaking relationships in a fluid community membership.

1 Model

We present a truly dynamic statistical generative network model that captures membership, formation, and fluidity of community membership and the resulting structure of interactions. This model incorporates some of the most fundamental properties of the real–world networks. A time evolving network generative model that evaluate network community structure is presented in [6]. However, the multiplicity of scale and relationships renders it harder to analyze in detail.

Table 1: Parameters for the generative model

Parameter	Description
N	Size of the network
T	Number of timesteps
$\langle C_0, \ldots, C_m \rangle$	Distribution of community sizes.
P_{intra}	Intra–cluster link probability.
P_{inter}	Inter-cluster link probability.
Optional parameters to setup the dynamics of link formation within communities. For example, for *preferential attachment model*, skewness and average degree.	

At a high level, the interactions in our dyanmic network generative model are driven by the individuals' membership in informal communities. Individuals tend to interact more within a community than accross communities however, over time they may update their affiliations.

2 The dynamic generative process

The network generative process works as follows. The model requires the parameters listed in 1. Given the size of the network N and the number of timesteps T, in each timestep $t_i \in T$, we generate a set of communities according to the given community distibution vector $\langle C_0, \ldots, C_m \rangle$. Nodes $1, \ldots, N$ are proportionally assigned to each community. Then, based on the P_{intra} probability, nodes within a community are linked to each other. In the next step, the nodes are connected across communities using the probability P_{inter}. The above steps are repeated for each timestep $t \in T$. Once the community distribution and node affiliations are fixed in each timestep, nodes are switched from one community to another community across timesteps based on the probability P_{sw}. Note that this process of community affiliation over time roughly follows the stochastic block model generative process with the evolution of the network with time incoporporated into it. Other than the basic uniform probability of interactions within a community P_{intra}, the model can take optional parameters to specify other individual level interactions pattern. For example, in the current version, this generative model can take *Preferential Attachement* model as the process by which nodes interact among themselves within a community.

3 Sample networks

We study the dynamic evolution of networks over a wide range of parameter values. For instance, to generate networks that have at most one giant community and other communities that are trivially small, while maintaining a preferential attachment model of connectivity within a community, we sample thousands of networks by controlling for the relevant parameters. 2 shows the degree distribution of a class of networks in which the giant component encompasses 40–50 % of the network and all the rest of the

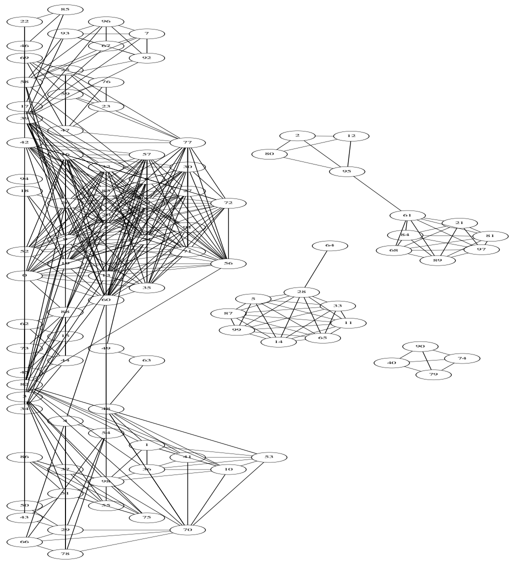

Figure 1: Modular structure of dynamic network

components are some constant fraction c of log of the network size N. For the preferential attachment model the exponent of skewness γ is set between 2 and 3, as has been shown for most real–world networks. For networks based on this model, the minimum average degree is set slightly above 1 to mimic the growth process. The resulting synthetic networks have a bimodal degree distribution. That is, there is a large frequency of smaller node degrees but also a relatively significant number of nodes that have degrees closer to the size of the largest component. This instantiation of the network generative process also captures the modular structure of the underlying dynamic network, as is evident by the sample graph in 1.

To verify how well this model imitates reality, we suggest two types of tests. First, measure the global network properties of the static representation of the dynamic sample network. Real–world networks have been shown to exhibit certain standard global properties, such as, most real–world networks belong to certain classes of degree distributions, have short geodesics, and/or high clustering coefficient. Comparing these real–world network properties with the ones exhibited by the dynamic network model based networks help us estimate the parameter settings of our model that correlate to the properties observed in real world. Secondly, using the maximum likelihood approach to measure the actual properties of the real–world dynamic networks. Such as, the probability of switching of individuals within communities, the probability of links within and across communities, the expected number of communities given an observed number of observations in time, and the sizes of the communities relative to the size of the population. These parameters can be estimated using the dynamic community detection model proposed by Chayant et al. [8]. We can use the parameters estimated by the

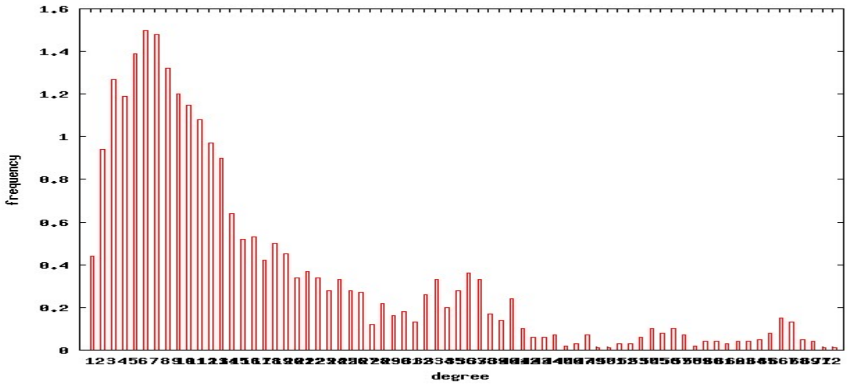

Figure 2: Degree distribution of skewed community structure

dynamic community detection model to synthetically generate similar networks using the generative model proposed here and measure how well the real–world dynamics of interactions and group formation can be replicated.

4 Conclusion

Research in computational generative modeling of networks has, over more than a decade, tried to build as realistic models as mathematically and statistically possible. However, for the most part they have failed to capture the complexity of multiplicity of properties exhibited by such networks. In this work, we propose a generative model for dynamic networks based on the notion of distribution of communities within the population that split and merge over time. Thus, this model not only emulates the dynamics of individual level interactions within communities but it also incorporates the structural changes of the communities themselves.

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.

[2] A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435:207–211, 2005.

[3] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.*, 99:8271–8276, 2002.

[4] Mark S. Handcock and Martina Morris. A curved exponential family model for complex networks. *Comput. Math. Organ. Theory*, 15(4), December 2009.

[5] Newman M. and Girvan M. Finding and evaluating community structure in networks. *Phys. Rev.*, 69, 2004.

[6] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. *Science*, 328(5980):876–878, May 2010.

[7] M. E. J. Newman. Random Graphs with Clustering. *Physical Review Letters*, 103(5):058701+, July 2009.

[8] Chayant Tantipathananandh, Tanya Berger-Wolf, and David Kempe. A framework for community identification in dynamic social networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, pages 717–726, New York, NY, USA, 2007. ACM.