Dynamic networks generative model for skewed component distribution

Habiba Chayant Tantipathananandh Tanya Berger-Wolf
hhhabib3@uic.edu ctanti2@uic.edu tanyabw@uic.edu
Department of Computer Science, University of Illinois 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 with 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 glaring shortcoming of this second class of models is that these models 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 upto a certain point in time.

Here is a simple example of what these models fail to capture. At given point in time, society exist as a collection of loosely formed communities [3]. As an individual joins a community, 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.

In this work we present a truly dynamic statistical generative network model captures membership, formation, and fluidity of community membership and the resulting structure of interactions. While not essentially complete, 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.

At a high level, the interactions in our model are driven by the individuals’ membership in informal communities. Individuals tend to interact more within a community than with others but can change their affiliations over time. The parameters of the model are listed and described in Table 1.

We studied 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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Size of the network</td>
</tr>
<tr>
<td>T</td>
<td>Number of timesteps</td>
</tr>
<tr>
<td>(C₀, ..., Cₘ)</td>
<td>Distribution of communities.</td>
</tr>
<tr>
<td>P_{intra}</td>
<td>Intra-cluster link probability.</td>
</tr>
<tr>
<td>P_{inter}</td>
<td>Inter-cluster link probability.</td>
</tr>
</tbody>
</table>

Optional parameters to setup the dynamics of link formation within communities. For example, for preferential attachment model, skewness and average degree.
that are trivially small, while maintaining a preferential attachment model of connectivity within a community, we sampled thousands of networks by controlling for the relevant parameters. Figure 1b shows the degree distribution of class of networks in which the giant component encompasses 40 to 50% of the network and all the rest of the components are some constant fraction of log of the network size. For a preferential attachment the exponent of skewness is set between 2 and 3 as is the case with most real world networks and minimum average degree is set slightly above 1 to mimic the growth process. We observe a bimodal degree distribution in such synthetic networks, where there is a large frequency of smaller degrees but also a relatively significant number of nodes have degrees closer to the size of the largest component. We also are able to capture the modular structure of the underlying dynamic network, as is evident by the sample graph in Figure 1a.

To verify how well this model imitates the reality, we test it in two ways. First, we measure the global network properties of the dynamic network model. Real-world networks have been shown to belong to certain classes of degree distributions, have short geodesics, and/or high clustering coefficient. We estimate the parameter settings of our model that correlate to the properties observed in real world. Secondly, we use 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. We use those parameters to synthetically generate similar networks using our generative model and measure how well we can replicate the real process.

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 paper, 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. This model not only mimic the dynamics of social individuals organizing themselves in communities but it also incorporates the changes that occur in such group structures over time.

References