Formation of Multiple Networks*

Matteo Magnani[1] and Luca Rossi[2]

[1] Dept. of Computer Science
Aarhus University, Denmark
magnanim@cs.au.dk
[2] Dept. of Communication Studies
University of Urbino, Italy
luca.rossi@uniurb.it

Abstract. While most research in Social Network Analysis has focused on single networks, the availability of complex on-line data about individuals and their mutual heterogenous connections has recently determined a renewed interest in multi-layer network analysis. To the best of our knowledge, in this paper we introduce the first network formation model for multiple networks. Network formation models are among the most popular tools in traditional network studies, because of both their practical and theoretical impact. However, existing models are not sufficient to describe the generation of multiple networks. Our model, motivated by an empirical analysis of real multi-layered network data, is a conservative extension of single-network models and emphasizes the additional level of complexity that we experience when we move from a single- to a more complete and realistic multi-network context.

Keywords: Multi-layer networks, Network formation, Social Network Sites.

1 Introduction

Network formation models are among the most important tools in Network Science and Social Network Analysis (SNA). A typical application of artificially generated networks is to provide *null* models that can be used to test new measures and make comparisons with real networks, so that significant patterns can be highlighted in the real data. In addition, these models are useful to test hypotheses on the dynamics underlying network evolution.

However, existing generative models have been developed to describe the evolution of *single* networks. While this is very relevant, as most of the research in SNA has been devoted to single networks, recent empirical studies have emphasized how on-line social systems including Social Network Sites (SNS) are made of multiple stratified networks influencing each other [1,2]. Multi-network

* This work has been supported in part by the Italian Ministry of Education, Universities and Research PRIN project *Relazioni sociali ed identità in Rete: vissuti e narrazioni degli italiani nei siti di social network* and FIRB project RBFR107725.

A.M. Greenberg, W.G. Kennedy, and N.D. Bos (Eds.): SBP 2013, LNCS 7812, pp. 257–264, 2013.

models were discussed several years ago in the field of SNA (also known as multiplex networks, [3]) and described as everyday experience by sociological research [4], but only recently the availability of real multi-network data has boosted the development of new models [5,6] and algorithmic approaches [7,8] based on the assumption that the analysis of the single networks may provide a distorted scenario if their multi-layered organization is not taken into consideration. As a simple example, on-line information propagation is typically characterized by the traversal of different networks [9].

In this paper we introduce the problem of multi-layer network generation. This is a challenging task, because models describing the formation of multiple networks should still generate network layers compatible with existing models and experimental observations of single networks, but should also consider the mutual relationships between different layers. Therefore, we propose a model where network evolution may be characterized both by *internal* dynamics, as described by existing single-network models like Preferential Attachment, and by *external* dynamics, where events like the creation of a new connection are influenced by the structure of other networks (here called *network layers*).

The paper is organized as follows. In the next two sections we briefly review the main theoretical basis of our work, namely network formation and multi-layer models. Then, in Section 4 we propose our approach. Our work is based on an analysis of real data that are used as a guideline for the definition of our model and also to test its ability to reproduce real observations. These data are presented in Section 5.

To the best of our knowledge our model is the first to deal with the generation of multi-network data. As such, it raises many new questions regarding the parameters and processes to be used to represent the dynamics underlying the formation of multiple networks. We devote our concluding remarks to these issues.

2 Network Formation Models: A Quick Review

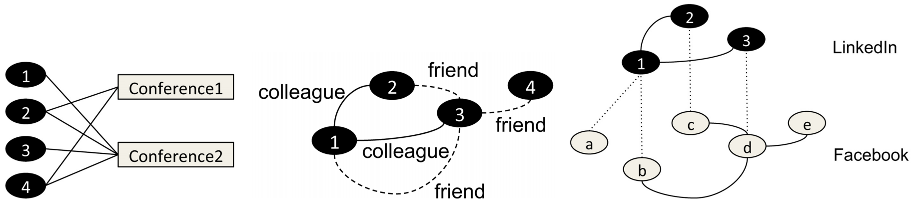
Research on random network models, their definition and related algorithms, is at least as old as modern network science and it has always been characterized by a common goal: being able to reproduce networks as they are observed in social, biological or physical phenomena. Within this perspective, we provide an essential summary of the most popular network models.

The definition of more and more sophisticated models can be seen as a never-ending attempt to catch the true complexity and inner nature of networks [10]. Among the first attempts in this direction, the Erdős-Rènyi model [11], often notated as the *G(n,p)* model, provides a simple but effective way to generate basic random networks. While this model has been historically useful to rise the interest on research topics such as edge probability and normal degree distribution, it fails at describing networks appearing in real-life phenomena. Its major caveats, i.e., the lack of scale-free degree distribution and the lack of high clustering values that are often observed in real-life contexts, have later been addressed by the Barabàsi-Albert and the Watts-Strogatz models [12,13]. The Barabàsi-Albert model is based on the concept known as Preferential Attachment, stating that important nodes in a network have higher probability than others to further increase their popularity. In addition, like in other more recents models [14,15] Preferential Attachment is not only used as a method to generate a network, but describes its formation step by step — in particular, the growing aspect is essential to obtain the required degree distribution.

While all these models provide a rather detailed level of description of several existing networks, and have been fruitfully used to simulate many real-world phenomena, none of them supports a multi-layer structure, therefore they cannot be directly used to describe the whole complexity of entangled multi-layered social phenomena.

3 Multi-layer Network Models

In the recent literature on multiple social network models we can find proposals allowing multiple node types [16,17], exemplified in Figure 1(a), models allowing multiple relationship types [5,8], represented in Figure 1(b), and models explicitly representing the co-existence of multiple networks (also called multi-layer networks) [6], represented in Figure 1(c).



(a) Multi-type (b) Multi-relationship (c) Multi-layer

Fig. 1. Three examples of heterogeneous networks: an author-conference graph (a), a multi-relationship network (b) and a multi-layer network (c), made of multiple social graphs and mappings indicating that different nodes correspond to the same individual

In the following we use a multi-layer network containing only nodes related to individuals (therefore, we do not consider heterogeneous nodes). In addition, we only allow nodes to have a single correspondence with nodes in other layers — the more complex situation of a node corresponding to n different nodes in another layer, e.g., a Facebook user having multiple Twitter accounts, is left to future extensions. As it appears from Figure 1(c), the main constituents of this model are two or more network layers, not dissimilar from traditional networks, and *mappings* indicating which nodes in different layers correspond to the same individual.

4 From Single- to Multi-layer Generators

Figure 2 shows a generic process of network formation. We call this network N_1. At different timestamps $t_0, t_1, \ldots$ a new node (+n) or a new edge (+e) are added to the network[1]. The specific mechanisms regulating the creation of nodes and edges vary depending on the formation model. In the following, we will adopt a well known model using Preferential Attachment to generate directed networks [14]. In summary, this model chooses the nodes to be connected together either at random, or with a probability proportional to the in- and out-degrees of the nodes.

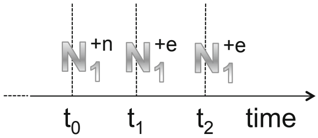


Fig. 2. Abstract view of the evolution of a network

Figure 3 extends the previous example to two networks N_1 and N_2. If we focus on a single horizontal layer, say N_1, we can observe the same dynamics of Figure 2. However, the whole process highlights two main new aspects. First, considering two or more networks we can no longer assume that at every timestamp t_i an event happens in all networks. Therefore, every network will have some associated probability of *no action*. This probability is useful to model the fact that different networks may grow at different speeds. The second fundamental aspect consists in the fact that an action on one network may be influenced by a previous action on another one. In our example, an edge is created in N_1 following the fact that the same two nodes were already linked in N_2. Practically speaking, if I already know someone, e.g., we are friends on Facebook, this may increase the probability that we will also connect on another on-line social network.

In summary, according to our model at every time t_i there are three possible events on each network:

1. **no-action**: nothing happens, i.e., the network remains unchanged.
2. **internal-growth**: the network grows according to internal dynamics, i.e., something happens independently of the other networks. For example, a Twitter user may find a tweet interesting and thus start following its author. In the following this event will be modeled as a Preferential Attachment process.

[1] As in all the aforementioned models, in this paper we only consider growing networks and not the deletion of nodes and edges, to keep this initial model simple and focus on the multi-layered aspects of network evolution. The extension of the model to deletion events will be object of future work.

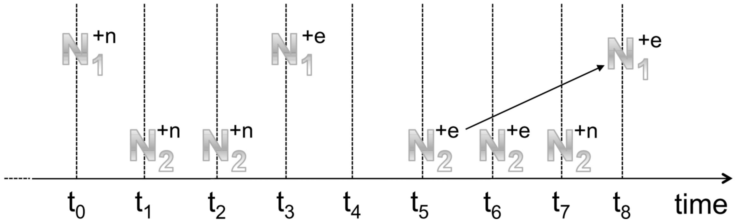


Fig. 3. Abstract view of the joint evolution of two networks

3. **external-growth**: the network grows according to external dynamics, i.e., something happens because of the configuration of an external network. For example, Dante and Beatrix are already "off-line" friends and after opening their accounts they also start following each other on Twitter (it is worth noting how in this example we do not limit our model to on-line social networks).

As said, in the following we will use Preferential Attachment as a network formation model in case of internal growth [14]. However, nothing prevents us to consider other models that can be seamlessly plugged into our approach — we do not further develop this idea because of space limitations and also to provide a well-defined first version of our model.

On the contrary, we need to discuss more the event of external growth. In this case, we can either add an edge coming from an other network, or a node. The first important aspect is that different networks may be more or less correlated, therefore the probability that the edge or node is "imported" from a specific other network is not uniform. The second important point regards the choice of the node to be imported and the corresponding creation of a link. We allow two possible actions:

1. The new node is just imported from the other network *at random*, meaning that the choice is not influenced by other nodes already in the target network[2].
2. The new node is chosen from the set of nodes connected to individuals already in the target network.

In practice, these actions can be exemplified as follows. Guido has an account on Twitter and an account on Facebook. At some point, another Facebook user creates an account on Twitter. Under option (1), this is just a random user that decided to join Twitter. Under option (2), this is a friend of Guido on Facebook who decided to join Twitter and start following Guido.

[2] By *target network* we indicate the network into which we are inserting the new node or edge.

5 Experimental Analysis

Our experimental analysis has two main objectives. The first is to highlight the presence of the theoretical features of our model in real multi-network data. The second objective is to test the ability of the model to replicate these data.

The data used in our experimental analysis of a real multi-layer network has been initially extracted from Friendfeed, a social media aggregator [18]. In this system while users can directly post messages and comment on other messages much like in Facebook and other similar SNSs, they can also register their accounts on other systems. In this way, using the Friendfeed API we could retrieve the multiple accounts of the same users for several social services.

As a result we obtained a Friendfeed network with 7 677 120 arcs, a Twitter network with 37 805 211 arcs and a YouTube network with 708 911 arcs. These networks have been used for the analysis of degree centrality correlations reported in the following. In addition, we also built three networks by keeping only those connections between users in our sample, with respectively 37 997, 67 123 and 1 185 arcs. The (not surprising) different sizes of these networks motivate the *no-action* steps in our network model.

The left hand side of Figure 4 shows the correlation between user rankings according to their degree centrality on the Twitter network and on the Friendfeed network, while on the right of Figure 4 we have shown the correlation between user rankings according to their degree centrality on the Twitter network and on the YouTube network. To interpret these figures consider that each point represents a user, and users with a high x or y coordinate are among the top users on the corresponding SNS according to their degree centrality (more precisely, x and y coordinates correspond to the ranking of the user, 0 for the user with lower degree centrality, up to 7 628 for the user with the highest degree centrality in that SNS).

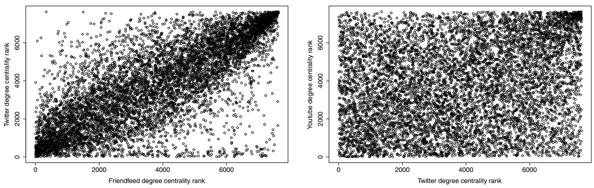
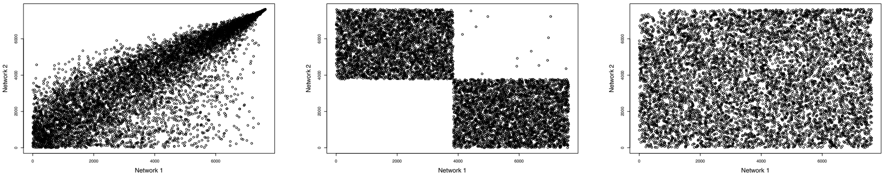


Fig. 4. User ranking (according to their degree centrality) in different networks: Friendfeed and Twitter (left) and Twitter and YouTube (right): Pearson correlation indexes are respectively .75 and .21

These figures show an interesting phenomenon corresponding to the varying probability of pairs of networks to be correlated that can be found in our network formation model. The high correlation between Friendfeed and Twitter means that users with a high degree centrality on Twitter tend to maintain it on the

Friendfeed network. On the opposite side, when we compare the degree centrality ranking on Twitter with the one on YouTube we are unable to detect a clear linear relationship.



(a) Correlated networks (b) Uncorrelated networks (c) Uncorrelated networks (same users)

Fig. 5. Degree rank correlation of network pairs generated using our model

In Figure 5 we have represented the corresponding rankings computed and compared on artificial networks generated using our model. In particular, Figures 5(a) and 5(b) correspond respectively to a pair of correlated networks and a pair of uncorrelated networks. While Figure 5(a) shows a behavior similar to the one observed on the Friendfeed and Twitter networks (Figure 4, left), the uncorrelated pair of networks (Figure 5(b)) presents a peculiar distribution of the degree rankings not observed in the real data. However, this can be explained by noticing that the way in which we collected our real data introduced a bias as we only selected individuals present in all the networks. On the contrary, our model does not enforce this controlled choice of users, and the uncorrelated growth of two networks results in different users joining either one network or the other, producing the well separated plot in Figure 5(b). We can make the hypothesis that this is what we would observe by comparing two unrelated real social networks, e.g., the QZone and Cloob networks[3]. In fact, by simply adding the constraint that users should be selected from a common basis, our model finally produces the networks corresponding to Figure 5(c).

6 Concluding Remarks

Multi-layer network data are everywhere in on-line social networks, but due to legal, privacy-related and technical issues they are still very hard to collect. This is also why research on topics such as the definition of new centrality metrics for these networks or the study of propagation patterns in multi-layer contexts is still at its early stages, although having been marked as very relevant for many years (e.g., in [3]). Therefore, the availability of our model could boost this area of research providing a tool to generate prototypical ML-networks for experimental researches [19]. At the same time, as far as new real data are collected our model and its variations can be used to test hypotheses regarding the evolution of multiple correlated networks.

[3] Respectively, the principal SNSs in China and Iran.

In our opinion, while it is still difficult to provide a thorough experimental analysis of our approach because of the limited availability of real data and the novelty of the topic, this paper draws many new research questions. A certainly non-comprehensive list includes the study of models for multiple networks where nodes and edges can be deleted, where a node in a network may correspond to multiple nodes in another, and where different networks evolve according to different internal formation models.

References

1. Szell, M., Stefan Thurner, R.L.: Multirelational organization of large-scale social networks in an online world. PNAS (107), 746–759 (2010)
2. Rossi, L., Magnani, M.: Conversation practices and network structure in twitter. In: ICWSM (2012)
3. Wasserman, S., Faust, K.: Social Network Analysis. Cambridge University Press (1994)
4. Goffman, E.: Frame analysis: an essay on the organization of experience. Harper & Row, New York (1974)
5. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., Pedreschi, D.: Foundations of Multidimensional Network Analysis. In: ASONAM, pp. 485–489 (2011)
6. Magnani, M., Rossi, L.: The ML-model for multi-layer social networks. In: ASONAM, pp. 5–12. IEEE Computer Society (2011)
7. Rodriguez, M.A., Shinavier, J.: Exposing multi-relational networks to single-relational network analysis algorithms. J. Informetrics 4(1), 29–41 (2010)
8. Brodka, P., Stawiak, P., Kazienko, P.: Shortest Path Discovery in the Multi-layered Social Network. In: ASONAM, pp. 497–501. IEEE Computer Society (2011)
9. Magnani, M., Montesi, D., Rossi, L.: Information propagation analysis in a social network site. In: ASONAM, pp. 296–300. IEEE Computer Society (2010)
10. Newman, M.E.J.: Networks: an introduction. Oxford University Press (2010)
11. Erdős, P., Rényi, A.: On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutató Int. Közl 5, 17–61 (1960)
12. Barabási, A., Albert, R.: Emergence of scaling in random networks. Science 286(5439), 509–512 (1999)
13. Watts, D., Strogatz, S.: The small world problem. Collective Dynamics of Small-World Networks 393, 440–442 (1998)
14. Bollobás, B., Borgs, C., Chayes, J., Riordan, O.: Directed scale-free graphs. In: SODA, pp. 132–139 (2003)
15. Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., Upfal, E.: Stochastic models for the web graph. In: FOCS, pp. 57–65 (2000)
16. Cai, D., Shao, Z., He, X., Yan, X., Han, J.: Community Mining from Multi-relational Networks. In: Jorge, A.M., Torgo, L., Brazdil, P.B., Camacho, R., Gama, J. (eds.) PKDD 2005. LNCS (LNAI), vol. 3721, pp. 445–452. Springer, Heidelberg (2005)
17. Sun, Y., Han, J., Zhao, P., Yin, Z., Cheng, H., Wu, T.: RankClus. In: EDBT, p. 565 (2009)
18. Celli, F., Di Lascio, F.M.L., Magnani, M., Pacelli, B., Rossi, L.: Social Network Data and Practices: The Case of Friendfeed. In: Chai, S.-K., Salerno, J.J., Mabry, P.L. (eds.) SBP 2010. LNCS, vol. 6007, pp. 346–353. Springer, Heidelberg (2010)
19. Magnani, M., Rossi, L.: Pareto Distance for Multi-layer Network Analysis. In: Greenberg, A.M., Kennedy, W.G., Bos, N.D. (eds.) SBP 2013. LNCS, vol. 7812, pp. 249–256. Springer, Heidelberg (2013)