Revealing Response Mechanisms in Online Learning Networks: A Graph Mining Approach

Moshe Mazuz
The Open University of Israel
mazuzmoshe@gmail.com

Reuven Aviv[1]
Tel Hai Academic College, and
The Open University of Israel
reuvenaviv@gmail.ac.il

Abstract

The goal of this research is to identify response mechanisms in online learning networks. We ask whether actors choose their response partners at random or whether certain special mechanisms are at work. In that case, we would like to discover what mechanism is most descriptive of the networks. While previous studies checked a few selective attributes of the networks, in this research, we capture their rich complex feature space by mapping them into a high-dimensional feature space. A Multi-way Support Vector Machine algorithm is used to classify 35 observed response networks of online learners into a set of five representative stochastic network generation models. The result shows that all the response networks were classified into a *preferential response* model in which actors tend to respond to partners who are a-priori equipped with response attraction power. We provide a possible explanation for this behavior, based on the nature and goal of the online learning networks, and discuss ways in which the study can continue.

Keywords: Online learning networks, preferential responses, network generation, graph mining, classification.

Introduction

Crucial for collaboration in the learning community is the development of responsiveness among participants (Rafaeli, Sudweeks, Mabry, & Konstan, 1998). A network of responsiveness is created among members of the community. The qualities of these networks, and whether there are hidden mechanisms that direct actors to choose their response partners, are not fully understood. Several studies attempted to reveal the properties and evolution mechanisms of these networks, examined their particular features and suggested models that fit the data (Aviv, Erlich, & Ravid, 2007a, 2007b; Cho, Stefanone, & Gay, 2002; Haythornthwaite, Kazmer, Robins, & Shoemaker, 2000; Martinez et al., 2002; Reffay & Chanier, 2002). One should note, however, that the models suggested focused on specific features of the networks and hence did not describe the actual network generation mechanism. Several network generation mechanisms can generate different networks with the same set of features. In general, a small number of network features are insufficient to decide which of the possible network generation models best describes the observed network.

This difficulty can be overcome by representing networks through high dimensional, theoretically infinite, feature vectors (Middendorf et al., 2004). An observed network is a point in the feature space. On the other hand, a stochastic network generation model creates an ensemble of networks, represented by a "cloud" of points in the feature space. One can design a set of models; each of which generates an ensemble of networks represented by "clouds" of

[1] On sabbatical at King Mongkut's University of Technology, North Bangkok (KMUTNB), Thailand

Proceedings of the Chais conference on instructional technologies research 2009: Learning in the technological era
Y. Eshet-Alkalai, A. Caspi, S. Eden, N. Geri, Y. Yair (Eds.), Raanana: The Open University of Israel

points in the feature space. These "clouds" can be used by graph mining tools to create classifiers of high accuracy. Therefore, associating an observed network with the best network generation model becomes a classification problem. The classifiers are used to find the network generation model that best describes the observed networks (Tan, Steinbach, & Kumar, 2006). This is what we did in this research.

The outline of this paper is as follows: formulation of the research goal is described in the second section. In the two following sections, we present the database and the candidate network generation models used in this study, respectively. In the fifth section we describe the method of analysis, and in the sixth section, we present the results. Discussion and directions for future research are in the last section.

Research Goal

Our goal was to find out whether a set of observed response networks of online learners can be classified into few classes (hopefully one), where each class was generated by a network generation model, selected from the literature, that has a different mechanism. This means that the structure of networks in a class fit the predictions of the associated network generation model substantially better than predictions of the other candidates. (The accuracy of this fit will be defined in the Methodology section). Identifying the network generation model can thus lead to better understanding of the behavior of actors in online learning networks.

The Open University of Israel (OUI) is a distance learning institution, based heavily on intensive use of learning technologies, with optional face-to-face tutorials. Each course utilizes at least one online learning network, usually over one semester (16 weeks). The objectives of the networks vary – from collaborative knowledge construction, to social, pedagogical or technical support. Objectives are not mutually exclusive. Accordingly, size, response links and participation patterns vary from course to course and from semester to semester. Numbers of participants in the network vary from 10 to 150, but most have about 50 students.

The Database

In this study, we selected for analysis a sample of 35 of about 500 online distance-learning networks in the OUI. Networks included in the sample were selected at random, omitting five networks in which the number of active participants (those who post at least one message during the semester) was below an arbitrarily selected threshold of 10.

For each of these networks, we created its (observed) response network from the log files of the online communication. This is a directed binary network, in which nodes represent the actors (participants), and a directed edge from node i to node j means that actor i responded to posts from actor j during the semester. The *responsiveness* of an actor i is the outgoing degree of node i. For each response network, we construct its adjacency matrix A: rows and columns are labeled by the nodes; an entry (i, j) equals 1 if there is an edge from i to j.

Network Generation Models

Numerous studies on complex networks have recently been performed. For a comprehensive review, see Albert & Barabasi (2002) and Newman (2003). The types of relations (edges) in these networks are very different, but it was found that many networks, large and small, have certain common features. Hence, a large group of generic network generation models tailored to the creation of the common features, were suggested (Albert & Barabasi, 2002; Middendorf et

al., 2004). From this group, we selected five model candidates, each of which is representative of a family of models:

Random Response (RR) Model

This model is the classic Directed Random Graph model (Erdos & Renyi, 1960). We assume that edges (response links) are created *at random*, with no special mechanism imposed. The number of nodes, and the probability of an edge from one node to another, are fixed.

Preferential Response (PR) Model

This model represents the "preferential attachment" family of models (Goh, Khang, & Kim, 2001), suggested to describe the behavior of many large, static networks, where the number of nodes (actors) is fixed. Nodes are labeled, where a node labeled **x** is assigned two *fixed a-priori probabilities*, $p_{x\text{-}in}$ and $p_{x\text{-}out}$, for incoming and outgoing edges, respectively. In our context, the important characteristics of this model are that the values of responsiveness of actors are distributed in a wide range, and some actors have relatively large values. This is the result of the a-priori probabilities for responsiveness $p_{x\text{-}in}$, $p_{x\text{-}out}$.

Dynamic Preferential Response (DPR) model

This model represents the "preferential attachment" family of models for *growing* networks (Krapivsky, Rodgers, & Redner, 2001). It was suggested for the generation of the World Wide Web. The number of nodes increases with nodes preferring to attach to other nodes that are already attachment-rich. Thus, at run time, each node learns the degrees of all others. Like the PR model, the values of responsiveness of the actors are distributed in a wide range, but this is the result of the *dynamic preference* of the actors to respond to response-rich actors during the growth process of the network.

Dynamic Copying (DC) Model

This model represents the "duplication" family of models (Vazquez, 2002), suggested for the description of citation and biological networks. Each node has only limited information about his neighbors, but gets additional information from them. At every step, a new node is added to the network, with an edge to an existing node selected at random. The added node chooses to connect to some of the neighbors of this neighbor. This last step is repeated recursively until no new neighbors are found. Like the PR and the DPR models, the values of responsiveness of the actors are widely distributed, but this is the result of the dynamic preference of the actors to respond to actors, based on the *information gathered* from other nodes during network growth.

Small World (SW) model

This model represents the family of "small world" models (Watts, 1999), explaining the "small-world" phenomenon in directed networks. The network is based on an underlying regular lattice, where each node has a fixed number of edges to its neighbors. The lattice network has high local clustering and long shortest-paths between nodes. To model a real network, edges between neighbors are randomly redirected to other, non-neighbor nodes, thereby creating shortcuts, which reduce the shortest paths without changing the local clustering much. The important characteristic in our context is that actors tend to respond within small fully connected sub-communities, with occasional responses to members in other sub-communities.

Methodology

A given network is represented by its NxN adjacency matrix **A**. Following Middendorf et al. (2004), we transform it into a multi-dimensional feature vector of length L with integer valued

coordinates. A coordinate counts the abundance of a certain feature of the network. In this study, we used 4680 features.

Each observed network is represented by a point in the feature space. The ensemble of networks created by simulating a network generation model is represented by a class of points in this space. Classifying the observed network into a network class is done by identifying the class of points nearest to the observed network point. This was done through support vector machine classifiers (Vapnik, 1995). Each pair-wise classifier classifies the tested network into one of two network generation models. For five network generation models, we needed 10 classifiers. These were built by the freely available C-implementation of SVM-Light (Joachims, 1999).

Each classifier was trained on 1/5 of the ensemble and then tested on the rest, repeated five times for cross validation. For a classifier C and an observed network T, the fraction of the number of errors obtained in training is the *training loss*, $\mathbf{L}_{C,T}$. The *accuracy* of the classifier **C**, $\boldsymbol{a}_{C,T}$, is the fraction of the number of correct classifications in the test phase. After testing, we classified the observed network into a winning model. If the winner is indeed close to the real (unknown) model in any aspect, it should be robust – it should not be sensitive to the particular set of features it uses. Hence, we randomly selected 30 different subsets of features, each containing 500 features, and then repeated the training, testing, and classification. The *robustness* of a classifier C with respect to observed network T, $R_{C,T}$, is the fraction of subset classifications that produced the winner. The final winning model for the observed network, T, was determined by counting votes for the different winners, among the pair-wise classifiers, using only the classifiers with robustness of at least 90%.

This was repeated for each observed network. Each network was classified into a final winner model. The quality of this classification is characterized by its training-loss, $\mathbf{L}_T$, accuracy, $\mathbf{a}_T$, and robustness, R_T, which are $L_{C,T}$, $\boldsymbol{a}_{C,T}$, and $R_{C,T}$ averaged on the 10 pair-wise classifiers.

A classifier may be wrong. To assess its overall predictive power, we calculated the probability that the classifier will provide the correct answer for a fraction *f* of a set of networks. This is the *accuracy* of classifier C at the confidence level *f*, which we denote by $\boldsymbol{a}_C(f)$.

Results

The goal of this study was to discover the mechanism that governs the creation of online learning networks. We used multi dimensional representation of the networks in order to create highly accurate pair-wise classifiers. These classifiers were used to distinguish between five network generation models of different mechanisms. Histograms of training losses among the 35 networks, for three classifiers, are presented in Figure 1.

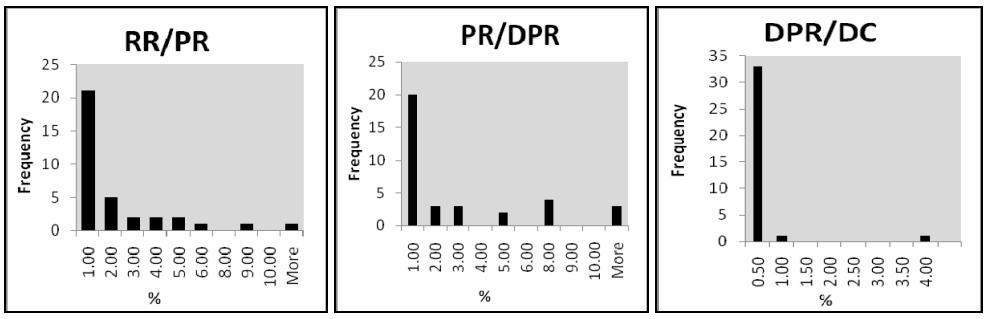

Figure 1: Histograms of Training Loss for three representative classifiers

Histograms of the accuracies among the 35 networks, for the same three classifiers are presented in Figure 2. The accuracy of all of the classifications for most networks, was above 90%.

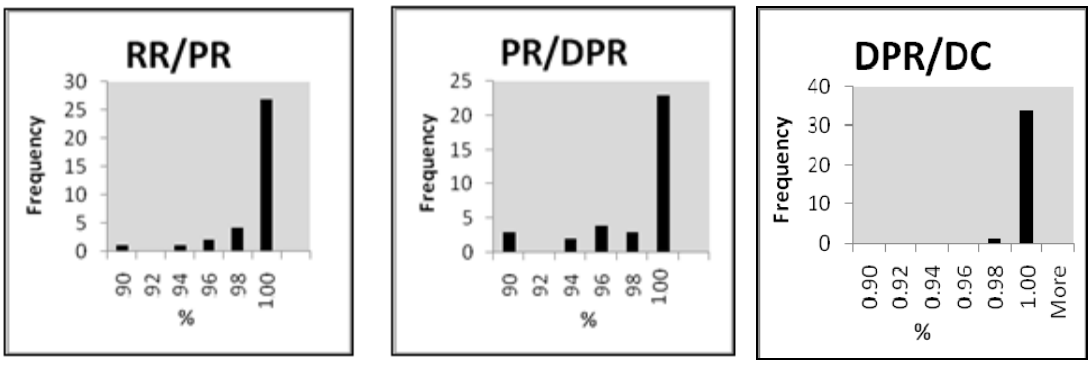

Figure 2: distributions of accuracies for three representative classifiers

In Table 1, we present the Training Losses, accuracies and robustness averaged over the 10 classifiers, the final votes, and the grand averages over all response networks.

Table 1: Summary statistics of the classification

Course Name	Training Loss L_T (in %)	Accuracy a_T (in %)	Robustness R_T (in %)	Votes				
				RR	PR	DPR	DC	SW
Algorithms2001b	0.34	99.7	91.6	2	4	3	1	0
Automata2001b	0.18	99.8	96.4	2	4	3	0	1
BasicConceptsIntRel2002a	0.60	99.6	93.8	2	4	3	1	0
CalculusII2002a	1.05	99.3	95.1	2	4	3	0	1
CompCoordinators2001a	1.83	98.4	96.0	2	4	1	0	3
CompCoordinators2002b	4.15	96.2	93.8	3	4	1	0	2
ComputerNetworks2001b	0.19	99.8	96.9	2	4	3	1	0
DataStructAndAlg2001b	0.00	100.0	97.3	2	4	3	0	1
DesignStudInCMC2002a	1.94	98.4	88.9	1	4	2	0	3
DiscreteMath2001a	0.00	100.0	96.9	2	4	3	0	1
EthicsB2000b	0.29	99.8	89.3	2	4	3	0	1
EthicsInBus2002b	1.12	99.0	95.1	2	4	3	1	0
GeneralBiology2001b	0.91	99.3	90.7	2	4	3	0	1
GuidingCapabilities2001a	1.62	98.8	96.0	2	4	1	0	3
Infi1-2001a	0.01	100.0	95.1	2	4	3	0	1
Infi1-2001b	0.01	100.0	88.4	2	4	3	0	1
Infi1-2002a	0.00	100.0	95.1	2	4	3	0	1
Informatics2001a	2.17	98.1	94.7	2	4	1	0	3
IntrCompSciPascal2002a	0.01	100.0	91.6	3	4	2	0	1
IntroPsychology2001a	0.06	100.0	97.8	2	4	3	0	1
LearningViaCMC2001a	1.48	99.0	95.1	3	4	1	0	2
LearningViaCMC2001c	4.02	96.4	89.8	3	4	2	0	1
LearningViaCMC2002a	3.94	96.4	94.2	3	4	1	0	2
OperatingSystems2001a	0.00	100.0	99.6	2	4	3	0	1
OperatingSystems2002a	0.00	100.0	94.7	3	4	2	0	1

Course Name	Training Loss L_T (in %)	Accuracy a_T (in %)	Robustness R_T (in %)	Votes				
				RR	PR	DPR	DC	SW
OrganizInfoSci2002a	1.30	98.9	96.9	2	4	3	0	1
PathwaysInChemistry2002a	0.66	99.5	95.6	2	4	3	0	1
PsychologicalTests2002a	0.05	99.9	95.1	2	4	3	0	1
Psychopathology2002a	0.29	99.8	95.6	2	4	3	0	1
PsychoPhysiology2001c	0.08	99.9	96.0	2	4	3	1	0
PsychoTests2002a	0.29	99.8	91.1	2	4	3	0	1
SoftwEngADA2001b	0.10	100.0	88.4	2	4	3	0	1
SoftwEngAda2002	0.05	99.9	90.7	2	4	3	0	1
WritingCMCWorks2001a	1.61	98.7	92.4	2	4	3	0	1
WritingWorksInCMC22002	0.62	99.5	97.8	2	4	3	0	1
AVERAGE ALL NETS	**0.88**	**99.0**	**94.0**	**2.14**	**4**	**2.54**	**0.14**	**1.17**

Training losses are of the order of a few percent, accuracies are above 90%, and robustness is 90% or higher. A vote is the number of times a particular Network Generation model won the classifications against the other four models, so its maximum value is 4. The winning model is clearly the Preferential Responsiveness (PR) model; all of the networks were classified into this model, with almost perfect robustness. The runner up, far behind, is the Dynamic Preferential Response (DRP) model. The Dynamic Copying (DC) and Small World models got very few votes.

Figure 3 presents the accuracies of the four classifiers as a function of confidence level *f*. The accuracies are above 90% for confidence level 0.92. That means that the probability that the classifiers will classify a set of networks correctly is above 90%. Some classifiers do even better.

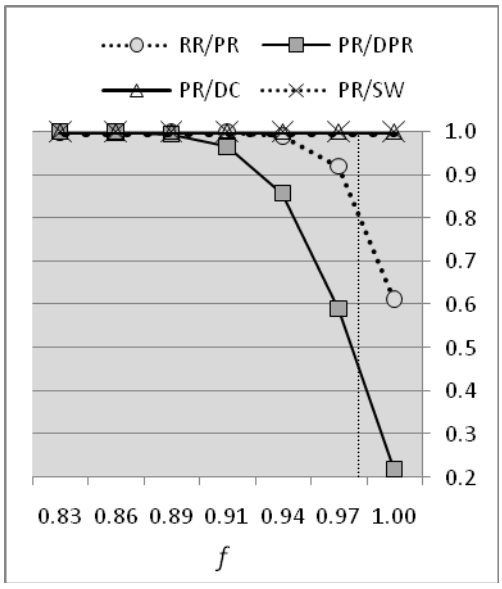

Figure 3: Accuracies of four classifiers as a function of confidence levels.

Discussion

It is no surprise that the Random Response (RR) model is not an adequate mechanism for the creation of response networks of online learners. This is not unlike almost any natural network that was analyzed (Newman, 2003). This was established before for network of online learners by analyzing specific features of the networks (Aviv et al., 2007b). In this study, we strengthen this conclusion by analyzing the multi-dimensional feature representation of the networks.

The predictions of the Preferential Responsiveness (PR) model fit almost perfectly with data. This model assumes that there were *a-priori* dis-homogeneous assignments of fixed probabilities to create and attract responses to each of the members in the online learning group. The result is in-homogenous distribution of the power of attracting responses. The assignments are presumably determined by attributes of the actors. Thus, positions of the actors in the network are important for creation or attracting responses, but they are not determined in the course of creating or developing the network.

This is in contrast to the Dynamic Preferential Response (DPR) model, which belongs to the preferential attachment family of models; these models are very successful in describing growing networks. In the DPR model, actors learn the network environment; they "look around" searching for the currently most powerful response creators and attractors. It seems that such learning of the environment process does not take place in online learning networks.

The Dynamic Copying (DC) model is also based on nodes learning the environment. In this model, actors look for neighborhoods of other nodes. This model is quite successful in describing biological and technical networks, but it seems that this learning mechanism does not take place in the development of online learning networks.

The Small World model is characterized by relatively high local clustering. In our context, this means that actors tend to respond to each other within small cliques. This does not happen. This confirms a previous analysis (Aviv et al., 2007a).

The a-priori winning assignments assumed by the PR model are presumably determined by attributes of the actors. In fact, it was shown (Aviv, Erlich, Ravid, & Geva, 2003) in one case, that a structured design of a network of online learners led to embedding of triggering and responsiveness into a certain set of students. These students took on bridging and triggering roles, without which the operation might have led to split groups or gaps in the construction schedule. This role taking significantly shifted the triggering and response power distribution from the tutor to some of the students, creating an in-homogeneous distribution of the response power.

The research described here has at least two obvious limitations. The first is the limited number of models. Although the models analyzed here are representative of families, it is still possible that other models would fit better, although the fit (measured by the accuracy) is already very good. A more severe limitation is the assumption that response links are binary. One can assign weights to the links to capture the frequencies of responses between actors. This issue is analyzed elsewhere.

All five classifiers show high values of accuracy and robustness. This implies that the conditional probabilities for the abundance of different features given a model would have similar behavior – they will all have a maximal value for the same model. This suggests that we might not need all the features captured by the feature vector to distinguish between models, as they all push toward the same model. To estimate this push power of a feature, we need to calculate the conditional probability of assigning a network to a model, given that the feature has a certain value (abundance). This intriguing idea will be pursued in future research.

References

Albert, R., & Barabasi, A. L. (2002). Statistical Mechanics of Complex Networks. *Review of Modern Physics, 74*, 47-97.

Aviv, R., Erlich, Z., & Ravid, G. (2007a). Analysis of Reciprocity and Transitivity in Online Collaboration Networks. *Connections*.

Aviv, R., Erlich, Z., & Ravid, G. (2007b). *Randomness and Clustering of Responses in Online Learning Networks*. Paper presented at the IASTED International Conference on Communication, Internet and Information Technology.

Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2003). Network Analysis of Knowledge Construction in Asynchronous Learning Networks. *Journal of Asynchronous Networks, 7*(3), 1-23.

Cho, H., Stefanone, M., & Gay, G. (2002). Social Network Analysis of Information Sharing Networks in a CSCL Community. In G. Stahl (Ed.), *Proceedings of Computer Support for Collaborative Learning (CSCL) 2002 Conference* (pp. 43-50). Mahwah, NJ: Lawrence Erlbaum.

Erdos, P., & Renyi, A. (1960). On the Evolution of Random Graphs. *Publ. Math. Inst. Hung. Acad. Sci., 5*, 17–61.

Goh, K.-I., Khang, B., & Kim, D. (2001). Universal behavior of load distribution in scale free networks. *Physical Review Letters, 87*(27).

Haythornthwaite, C., Kazmer, M., Robins, J., & Shoemaker, S. (2000). Community Development Among Distance Learners: Temporal and Technological Dimensions. *Journal of Computer Mediated Communication, 6*(1).

Joachims, T. (1999). Making large-scale SVM Learning Practical. . In B. Schlopf, C. Burges & A. Smola (Eds.), *Advances in Kernel Methods - Support Vector Machines*: MIT-Press.

Krapivsky, P. L., Rodgers, G. J., & Redner, S. (2001). Degree distributions of growing networks. *Physical Review Letters., 86*(23), 5401-5404.

Martinez, A., Dimitriadis, Y., Rubia, B., Gomez, E., Garrachon, L., & Marcos, J. A. (2002). Studying Social Aspects of Computer-Supported Collaboration with a Mixed Evaluation Approach. In G. Stahl (Ed.), *Proceedings of Computer Support for Collaborative Learning (CSCL 2002) Conference* (pp. 631-632). Mahwah, NJ: Lawrence Erlbaum.

Middendorf, M., Ziv, E., Adams, C., Hom, J., Koytcheff, R., Levovitz, C., et al. (2004). Discriminative Topological Features Reveal Biological Network Mechanisms. *BMC Bioinformatics, 5*, 181.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review, 45*(2), 167-256.

Rafaeli, S., Sudweeks, F., Mabry, E., & Konstan, J. (1998). ProjectH: A Collaborative Qualitative Study of Computer-Mediated Communication. In F. Sudweeks, M. L. McLaughlin & S. Rafaeli (Eds.), *Network and Netplay: Virtual Groups on the Internet* (pp. 265-282). Menlo Park, CA: MIT Press.

Reffay, C., & Chanier, T. (2002). Social Network Analysis Used for Modeling Collaboration in Distance Learning Groups. In S. A. Cerri, G. Guarderes & F. Paraguaco (Eds.), *Lecture Notes in Computer Science (LNCS)* (pp. 31-40).

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Boston: Pearson Education.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. NY: Springer-Verlag.

Vazquez, A. (2002). Knowing a network by walking on it: Emergence of scaling. *arXiv.org:cond-mat/0006132*.

Watts, D. J. (1999). *Small worlds: The dynamics of networks between order and randomness*. Princeton, NJ: Princeton Univ. Press.