



IMPROVING ALUMNI NETWORK EFFICIENCY WITH MACHINE LEARNING

V. Markovska^{1*}, S. Kabaivanov²

¹University of Food Technologies, Plovdiv, Bulgaria

²University of Plovdiv “Paisii Hilendarski”, Plovdiv, Bulgaria

ABSTRACT

Universities tend to rely on alumni organizations for providing long-term support for the institution and keeping track of successful students. Powered by network connectivity and recent advance in the area of social network research it's custom to have these organizations operating mostly in internet thus evolving into a dedicated small social networks. In this paper we analyze how machine learning can be used to improve efficiency of alumni organizations. SVM and K-means clustering algorithms are examined as tools that can not only give valuable information about members of the organization but also improve the level of support provided to the university by targeting the most appropriate alumni members for a particular event or project. Sample implementation is provided with demonstration on information from sample alumni networks.

Key words: support vector machines, machine learning, social network clustering

Alumni networks have slowly been making their way to Bulgarian universities but their success is still far from what has been seen in United States and Western Europe. While it is true that it takes time before new ideas and models of interaction are adopted and successfully applied, we believe that there are other reasons that are hindering efficient use of these networks. In this paper we argue what are the most important features of alumni networks that local universities need to concentrate on and suggest an efficient model to improve network efficiency with machine learning algorithms.

Why alumni networks are important?

There are many arguments why keeping touch with graduated students is important for universities but some of them are more relevant than others and need to be carefully explained in order to use alumni networks up to their full potential. **First of all**, in contemporary educational process the learning is getting more and more a lifelong experience, rather than limited in time stay in school or university. Palmer in [1], argues that technological changes are breaking the traditional way of learning as much as possible at early stages in life and then reaping off as much benefit from that education later on. As a result the necessity of keeping up with changes in a particular domain require a continuous education that goes beyond formal university

BSc or MSc programs. In such environment it is vital for every educational institution to keep in touch with its former students and use them in an efficient way to promote itself. Yet we need to stress that this relation is bi-directional and the reverse process may also take place – failing to keep good connection to former students may have negative impact to the institution.

Second, contemporary education is going one step further than simple teaching as its benefits also include the ability to work in a competitive environment, get connected to people with similar interests and gain social experience that is not strictly part of the studying syllabus. Alumni networks in this context help former students that have gained good positions to “pave the way” for new graduates of the same institution. Of course this does not need to apply to every single case but alumni networks can provide both useful leads and faster access to information about manager’s background, skills and preferences that may be crucial for getting a particular job.

Third, alumni networks can be an efficient way to support new students by providing scholarships as well as for supporting charity events that aim at helping the education institution. Saltman in [2] has successfully argued on the importance of philanthropy and support received by educational organizations. We should note however, that despite

immediate positive financial effects, alumni networks should not be treated only as a “free cash source” and their other positive effects can outweigh the direct financial support for students and universities.

Fourth, educational institutions can greatly benefit from keeping in touch with former students by using their expertise and firsthand experience when making changes to their programs, introducing new courses and updating existing methods of teaching. Put differently – alumni networks can be considered as a very large source for expertise and consulting information that is much cheaper than conducting market research and hiring external consulting companies.

While it is important to keep in mind the major advantages of building a successful alumni networks, it is also essential to be able to estimate their positive effects in order to compare them with the costs of keeping networks functional and up to date. Since some of the positive effects are very hard to measure they can only be estimated indirectly. Cohen and Malloy in [3] conclude that mutual fund portfolio managers tend to invest more in “connected” stocks (e.g. stocks of companies that are run by executives from the same college/university) with observed returns getting higher the stronger the connection between executives. Do, Lee and Nguyen [4] on the other hand estimate that external executive networks strongly influence investment decisions and valuation leading to positive effects of a bit over 4%, measured in company value increase. Although it is hard to link these estimates to value the benefits for an educational institution it is clear that the positive effects are both significant and require time to be developed.

Provided that benefits from building and maintaining a functional alumni network has been subject to extensive study, it is still common for Bulgarian universities to misuse

these networks or not pay attention to them at all. Crisp, Brown and Hardcastle in [5] note that alumni are also sometimes undervalued in well-established business schools, but at the time of writing the problem is much more visible for Bulgarian educational institutions where networks are sometimes implemented “pro forma” or in a way that simply does not allow to ever utilize their full power. Based on our firsthand experience, we believe that significant contribution to this issue can be attributed to the following factors:

- pure misunderstanding and/or misinterpretation of alumni networks, often accompanied with focusing on their technical implementation, rather than on their important features;
- lack of interest in long-term support of alumni networks, which requires substantial investment of money and time, but is required in order to achieve sustainable results;
- failure to capture the right mix of social networking and community features that alumni networks typically provide.

The last reason is particularly important as it requires to understand distinctive features of both communities and social networks in order to find a combination of both that suits best a particular educational institution. This also implies that alumni networks, although bearing the same name are not exactly the same across different universities and schools and even when considering the same institution the requirements may change over time. Differences between social networks and communities have been subject to different studies (as in [6], [7] and [8]) and a number of algorithms have been suggested to discover communities within existing networks (as in [9]) we are only interested in outlining those characteristics that require different approach toward building and maintaining a community or social network.

Table 1. Comparison of essential features and differences between communities and social networks

Social networks	Communities
Predefined relationships (for example links to colleagues or friends)	People are held together by common interest or shared values.
Connections are built gradually (one-at-a-time)	Connections can grow very quickly
In a typical case people have a limited number of network “memberships” (accounts if we speak technically).	It is normal to participate in multiple communities at the same time.
Network expansion grows over time, but the speed of growth is not restricted.	Communities take much longer time to develop and reach a mature state.

Table 1 highlights that alumni networks have features that can be distinctive for both social

networks and communities, thus they require to be analyzed with care and approaches that

treat them as pure social networks or pure communities are not going to fully utilize their advantages. However finding the right “mix” of special features requires extensive analysis and, what is even more important, needs to be repeated over time in order to keep our understanding up to date with changing environment. This also leads us to one crucial moment in creating a successful alumni network – continuous “maintenance” and all subsequent actions that aim to keep the network alive and active.

Alumni networks and machine learning

Automatic data analysis and model building can help a lot in improving efficiency of both new and existing alumni networks due to the following:

- it is a cheap method of continuous monitoring and analysis of the changes in the network, thus reducing the maintenance costs and making sure that fundamental changes can be detected in time;
- it can handle successfully huge volumes of data available in the alumni networks;
- it is a flexible set of tools that can be applied on the same data set for achieving different goals.

Not all machine learning algorithms are suitable for these tasks and there are multiple different approaches to address each of them. However we can split the major machine learning tasks, related to alumni networks into the following three groups:

- matching algorithms – responsible for calculating similarities between “nodes” within the network;

Matching algorithms are important in order to discover similarities (resp. calculate “distances”) between members of the alumni network and suggest links between them. Unlike standard social network relations these links are based on similarities defined through a matching criteria that can be either related to a specific project (for example charity project) or to a specific goal (for example finding proper apprentice positions for top performing students).

- searching algorithms – responsible for finding alumni that have specific characteristics;

Searching algorithms are quite straight-forward when it comes to finding alumni network members that match a certain search criteria. We have to note that despite being considered “standard solution” the usability of search algorithms depends on how accurate is the information stored in the network.

- clustering algorithms – responsible for

separating alumni network into different clusters;

Clustering algorithms are useful when it is necessary to target specific subgroups of the alumni network (like for example when conducting a specific support campaign). In such cases clustering can improve the efficiency by making sure that information/support requests are to reach the right alumni in the most suitable way.

Implementation of machine learning algorithms can ease maintenance of alumni networks, but it cannot resolve one of the major issues in the process – keeping the information up to date. We consider two generic approaches to address this problem:

- appointing responsible people (alternatively implementing algorithms) that check periodically available sources of information (like for example existing professional or social networks) and reflect changes in alumni network own database;

The main advantage of this approach is that data is safely stored within alumni network software reach and is easily accessible. However this comes at a cost, since one has to provide enough space to store all the information and there is no guarantee that changes will be reflected quickly (we assume that it is unlikely for alumni network members to take the effort to manually keep up to date yet another account in the alumni network).

- building alumni network database in a way that contains metadata about person links, interests and participation in existing networks without trying to replicate information that is already available elsewhere.

The main advantage of this approach is that only relevant information is stored in the network (e.g. results from machine learning algorithms) and for all “current” details like contact addresses, occupation details etc. etc. alumni network relies on information already available in other social networks. This also comes at a cost, that alumni network depends more on external sources of information.

Sample implementation and numerical results

Based on current efforts to build sample alumni network we have conducted several tests on efficiency of machine learning algorithms when used to provide metadata on former students. These tests used publicly available information from Twitter (thought information can be obtained from other networks like Facebook or LinkedIn) and the

last presidential campaign in USA. This particular data set has been used due to the fact that main topics discussed are already very

popular and accuracy of the algorithms can easily be assessed by professionals in different domains.

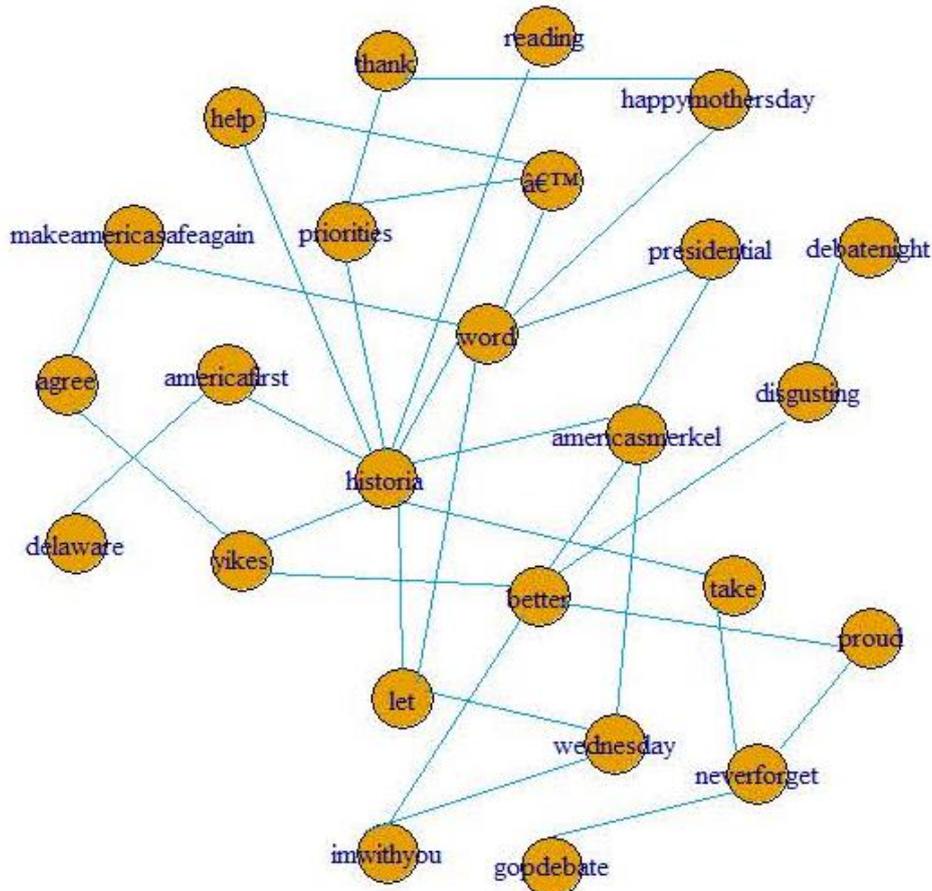


Figure 1. Network of frequently used terms

Our analysis is based on 6444 tweets¹ and uses their full text to search for major concepts and figure out if connections can be drawn between them. Text has been preprocessed to eliminate numbers, URL identifiers and special symbols. In addition text has been stemmed ([10]) using a dictionary of common English words. The preprocessed information has then been used to build a corpora and based on it to

calculate the most frequent terms and concepts (combination of terms), including the distances between them which we later use as a measure of association between terms and concepts. In order to use Euclidean distances the weights are first set in accordance with term/concept frequencies and then normalized. Thus a distance between two concepts x and y , each one existing in n dimensional space would be:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \tag{1}$$

The idea behind calculated distances is that they can be used to later on assess alumni interest in particular ideas and get them involved in specific projects, based on their own interests. As a result one can expect much higher participation rates as only alumni that

particular domain will be invited and involved. **Figure 1** shows part of the graph, including the top 25 terms from the analyzed text.

Preprocessed text can also be used to demonstrate how clustering algorithms can be applied to split alumni members in accordance to their “distance” toward a predefined criteria. We have applied K-means clustering ([11], [12]) in an attempt to separate tweets in ten different clusters, based on distance between used key terms.

¹Data can be accessed through Twitter or using one of www.kaggle.com datasets with the following R script: <http://www.kabaivanov.net/wp-content/uploads/2017/06/alumninetwork.r> have already expressed interest in the

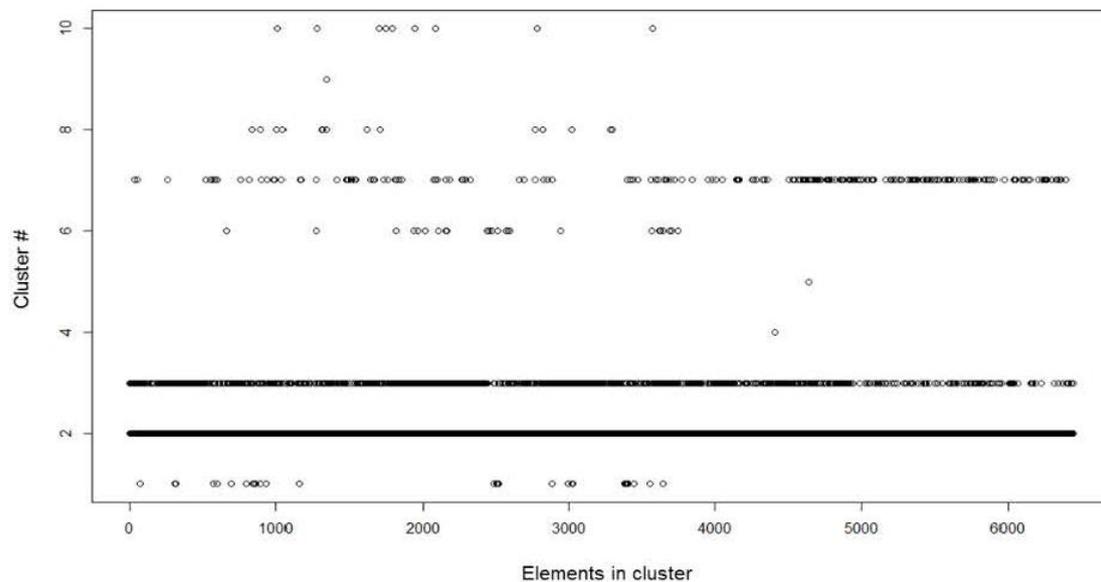


Figure 2. K-means clustering of used terms

With this particular data set, the results shown in **Figure 2**, demonstrate that there is a huge concentration of elements in clusters #2 and #3. This could be explained taking into account that typically political debates and campaigns are concentrated around a small number of key topics. In the context of alumni networks clustering can be used to analyze alumni based on specific criteria (for example how they are situated in the labor market but using their metadata and not statistical information for employment that may be outdated or not available at all).

CONCLUSIONS

Alumni networks have proven themselves as useful tool in the quest of building a competitive educational institution. In order to create a successful and sustainable alumni network it is necessary to take into account that they are much more than an “address book” or a simple mean to keep in touch with former students and require a continuous improvement.

We have argued that machine learning algorithms can help a lot in maintaining alumni networks and reduce greatly the costs of doing so. When alumni networks are designed from scratch with machine learning in mind, then this can also dramatically increase their value by making it possible to conduct different campaigns and keep former students involved with the development of their university or school. To further reduce costs and provide up-to-date information about alumni we stress that it is better to keep networks designed in a way that allows interaction with existing social and professional networks.

REFERENCES

1. Palmer, "Lifelong learning is becoming an economic imperative," *The Economist*, 2017.
2. K. Saltman, *The gift of education: Public education and venture philanthropy*, New York: Palgrave Macmillian, 2010.
3. L. H. Cohen and C. J. Malloy, "The Power of Alumni Networks," *Harvard Business Review*, vol. October, 2010.
4. Q.-A. Do, Y. T. Lee and B. D. Nguyen, "Directors as Connectors: The Impact of the External Networks of Directors on Firms," no. <https://ssrn.com/abstract=2753836>, February 2017.
5. Crisp, L. Brown and S. Hardcastle, "Alumni Matters 2012: What Alumni Want," Carringtoncrisp, London, 2012.
6. J. Scott, *Social network analysis*, Sage, 2017.
7. M. E. Newman, "Modularity and community structure in networks.," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577-8582, 2006.
8. G. Palla, A.-L. Barabási and T. Vicsek, "Quantifying social group evolution.," *Nature*, vol. 446.7136, pp. 664-667, 2007.
9. V. Poaka, S. Hartmann, H. Ma and D. Steinmetz, "A Link-Density-Based Algorithm for Finding Communities in Social Networks," *Advances in Conceptual Modeling*, vol. 9975, 2016.
10. Feinerer, "Analysis and Algorithms for Stemming Inversion," *Lecture Notes in Computer Science*, vol. 6458, pp. 290-299, 2010.
11. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm.," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100-108, 1979.
12. K. Jain, "Data clustering: 50 years beyond K-means.," *Pattern recognition letters*, vol. 31, no. 8, pp. 651-666, 2010.

