# Poster: Studying The Social Networks in Educational Forums (Summary)


- Mary Luz Mouronte-López.
- Higher Polytechnical School
- Universidad Francisco de Vitoria
- Madrid, Spain
- maryluz.mouronte@ufv.es



The Fifteenth International Multi-Conference on Computing in the Global Information Technology

ICCGI 2020

October 18, 2020 to October 22, 2020 - Porto, Portugal

# RESUME OF THE PRESENTER

- Ph.D. by Universidad Politécnica de Madrid in 2004. She has worked for more than 25 years in the telecommunication industry and now is Associated Professor at Universidad Francisco de Vitoria.
- Research fields: software systems that use intelligent agents, analysis and modelling of networks, study of complex systems of diverse nature such as telecommunications, social, technological, biological and medical.
- She has scientific contributions relating to telecommunication networks, intelligent agents and complex systems in scientific journals, national and International congresses. She holds several patents on algorithms applicable to telecommunication networks.

# GOALS

- Virtual Educational Platforms offer a rich set of tools, which are properly applied in teaching can be very useful in arousing the motivation of students and increase collaboration between teachers and students.
- Several research exists on Social Networks and Learning Management Systems [1][2][3].
- This research analyzes the social interactions that took place in Moodle, when this platform was used in the context of a university course. Several topological parameters and the structure of communities were calculated.

[1] Hassan, A. (2012). Social Network Based Learning Management System. IOSR Journal of Computer Engineering. 3. 18-23. 10.9790/0661-0321823.

[2] Cela, K. & Sicilia, M. & Sánchez, S. (2015). Social Network Analysis in E-Learning Environments: A Preliminary Systematic Review. Educational Psychology Review, 27. doi: 10.1007/s10648-014-9276-0.

[3] Tasneem A. et al. (2017). Learning Management System versus Social Networking Sites. International Business Research, 10 (6), 123-136.  doi: 10.5539/ibr.v10n6p123.

# ANALYZING THE SOCIAL NETWORKS

- The XML file of the Moodle forums was analyzed and processed using software programs implemented in Python.
- These programs were designed, built and tested, following the typical life cycle of any software component. The interactions in each forum were represented in a graph G = (V; S), where V is the set of nodes corresponding to students and faculties and S is the set of links between them.
- 14 forums each with an average of 115 students were analyzed. Three types of forums were considered: news and questions forums, practical exercise forums and theoretical content forums.

# ANALYZING THE SOCIAL NETWORKS (Cont.)

- Topological properties:
  - Betweenness centrality [4]  [6]
  - Node clustering coefficient [5]
  - Eigenvector centrality [5]
  - PageRank [5]
  - Similarities between vertices (Walktrap Algorithm [7] was used to identify communities)


[4] Newman, M. E.J. (2003).  “The structure and function of complex networks”, SIAM Review, 45, 167-256.

[5] Newman, M.E.J. (2002).  “Assortative Mixing in Networks”. Physical review letters, 89 (20), 48109–1120.

[6] Boccaletti et al. (2006). Complex networks: Structure and dynamics. Physics Reports, 424, 175 – 308

[7] Pons, Pascal & Latapy, Matthieu. (2006). Computing Communities in Large Networks Using Random Walks. J. Graph Algorithms Appl, 10, 191-218. doi: 10.7155/jgaa.00124.


# BETWEENNESS

- The betweenness bi of a node is the number of times that a node appears between the shortest paths of two other nodes and thereby quantifying the importance of a node [6], and is defined as:

$$b_i = \sum_{i \neq j} \frac{n_{jk}(i)}{n_{jk}}$$

Where njk is the number of shortest paths connecting j and k, while njk(i) is the number of shortest paths connecting j and k and passing through i.

# CLUSTERING COEFFICIENTS

Clustering coefficient $C_v$ of a node *v*:

$C_v$ = *|E(N(v))|/(*max possible number of links in *N(v)*)

Where *N(v)* the neighborhood of *v,* i.e., all nodes adjacent to *v*

$C_v$ can be viewed as the probability that two neighbors of *v* are connected. Thus $0 \leq C_v \leq 1$.

For nodev of degree *0* or *1,* by definition $C_v = 0$.

# EIGENVECTOR CENTRALITY

Eigenvector centrality of a node n Є G :

$$X_n = {}^{1}/_{\lambda} \sum_{j=1}^{j=N} x_j = {}^{1}/_{\lambda} \sum_{j=1}^{j=N} A_{ij} * x_j$$

Where:

$A_{ij}$ is element ij of the Adjacency Matrix, such as $A_{ij}$ = 1 if node i is attached to node j and 0 otherwise.

This equivalent to $A*X = \lambda*X$ where $\lambda$ is the largest eigenvalue associated with A and X is its associated eigenvector.

# PAGERANK

PageRank  of a node n Є G:

$$PR(n) = (1 - \alpha) + \alpha * \sum_{w \in V:\, w \rightarrow n} \frac{PR(w)}{K_{out}(w)}$$

Where:

$\alpha$, damping parameter, $\in [0, 1]$.

PR(w) is the PageRank of the node w which is linked to n.

# COMMUNITIES

## Walktrap Algorithm

- This method uses random walks on *G* to identify communities. At each step in the random walk, the walker is at a node and moves to another node chosen randomly and uniformly from its neighbors.
- The sequence of visited nodes is a Markov chain where the states are the nodes of *G*.
- A: Adjacency matrix of N x N, bidimensional representation of the relationships between stops, where $A_{ij} = 1$ when a connection between $v_i$ and $v_j$ exists and $A_{ij} = 0$ otherwise.
- D: diagonal matrix of the degrees $\Delta_i$; $D_{ii} = k_i$ and $D_{ij} = 0$ where $i \neq j$.
- At each step the transition probability from node $v_i$ to node $v_j$ is $P_{ij} = A_{ij}/k_i$ , it is an element of the transition matrix P for the random walk.

$$P = D^{-1} A$$

# COMMUNITIES (II)

## Walktrap Algorithm

- The process is driven by the powers of the matrix P:
  - The probability of going from i to j in a random walk of length t is $P_{ij}^{\ t}$.
  - We define an inter-node distance measure:

$$S_{ij} = \sqrt{\sum_{K=1}^{n} \frac{(Pik^{t} - P_{jk}^{t})^2}{K_k}}$$

  - We define the probability to go from community C to node j in t steps as:

$$P_{Cj}^{t} = \frac{1}{|C|} \sum_{i \in C} P_{ij}^{t}$$

  - We define the distance between two communities as:

$$S_{C1\,C2} = \sqrt{\sum_{K=1}^{n} \frac{(P_{C1k}^{t} - P_{C2k}^{t})^2}{K_k}}$$

  - We can also define the distance between a node *i and a community C:* $S_{iC}$

# COMMUNITIES (II)
# Walktrap Algorithm

We start from a partition $\rho 1 = \{ \{i\} , i \in V\}$ ; of the graph into n communities reduced to a single node. We first compute the distances between all adjacent nodes. Then this partition evolves by repeating the following operations. At each step k:

- choose two communities $C_1$ and $C_2$ in $\rho_k$ according to a criterion based on the distance between the communities.
- merge these two communities into a new community $C_3 = C_1 \cup C_2$, create the new partition: $\rho_{k+1} = (\rho_k \setminus \{ C1UC2\}) \cup C3$, and update the distances between communities (we will see later).
- after n -1 steps, the algorithm finishes. Each step defines a partition $\rho_k$ of the graph into communities.

# COMMUNITIES (III)

## Walktrap Algorithm

- The algorithm uses an begins with one partition for each node (|p| = n ) .
- we will only merge *adjacent communities (having at* least an edge between them).At each step k , two communities are chosen based on the minimization of the mean $\sigma_k$ of the squared distances between each node and its community.

$$s_k = \frac{1}{n} \sum_{Ci \in pk} \sum_{i \in Ci} s_{iCi}^2$$

Instead of directly calculating this quantity first we calculate the variations $\Delta \sigma (C1, C2)$

- So for each pair of adjacent communities C1, C2; we compute the variation that would be induced if we merge C1 and C2 into a new community C3 = C1 U C2.
- We can efficiently calculate these variations as

$$\Delta \sigma (C1, C2) = \frac{1}{n} \frac{|C1||C2|}{|C1| \quad |C2|} s_{C1C2}^2$$

# COMMUNITIES (IV)

## Walktrap Algorithm

- The community merge with the lowest $\Delta\ \sigma$ is performed and the process is repeated again updating the values of s and $\Delta\sigma$ then performing the next merge.
- After n-1 steps, we get one partition that includes all the nodes of the network $|\rho n| = \{N\}$. The algorithm creates a sequence of partitions $(\rho k)_{1 \leq k \leq n}$ .
- Finally, we use modularity to select the best partition of the network, calculating Q $\rho k$ for each partition and selecting the partition that maximizes modularity.

# COMMUNITIES (V)

## Walktrap Algorithm

- We define modularity Q as the fraction of links within communities minus the expected value of the same quantity for a random network.

$$Q = \frac{1}{2m} \sum_{ij} \{A_{ij} - \frac{K_i K_j}{2m}\} \delta_{CiCj}$$

- where the $\delta_{CiCj}$ function is 1 if $C_i = C_j$ and 0 otherwise, m is the number of links in the graph, and $k_i$, $k_j$ are the degrees of the nodes **i,j**. The sum of the term $k_i k_j/2m$ over all node pairs in a community represents the expected fraction of links within that community in an equivalent random network where node degree values are preserved.

# RESULTS

The Fifteenth International Multi-Conference on Computing in the Global Information Technology
ICCGI 2020
October 18, 2020 to October 22, 2020 - Porto, Portugal

TABLE I. In each forum, averagre mínimum distance between nodes <l>, average betweeness <b>, average PageRank <PR> (considering $\alpha$=0.85), Average EigenVector centrality <EV>, Average Degree <K> and Average Clustering <C> values. [8]

| | *F1* | *F2* | *F3* | *F4* | *F5* | *F6* | *F7* | *F8* | *F9* | *F10* | *F11* | *F12* | *F13* | *F14* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| <l> | 1.22 | 1.13 | 1.34 | 1.01 | 1.78 | 1.15 | 1.85 | 1.11 | 1.13 | 1.18 | 1.02 | 1.15 | 1.85 | 1.94 |
| <b> | 0.007 | 0.007 | 0.012 | 0.009 | 0.007 | 0.013 | 0.008 | 0.006 | 0.015 | 0.013 | 0.006 | 0.008 | 0.010 | 0.002 |
| <PR> | 0.0002 | 0.0005 | 0.0090 | 0.0031 | 0.0042 | 0.0096 | 0.0063 | 0.0036 | 0.0107 | 0.0114 | 0.0043 | 0.0072 | 0.0078 | 0.0017 |
| <EV> | 0.0018 | 0.0013 | 0.0017 | 0.0078 | 0.0067 | 0.0100 | 0.0088 | 0.0056 | 0.0013 | 0.0238 | 0.0054 | 0.0086 | 0.0095 | 0.0025 |
| <K> | 16.10 | 15.01 | 13.01 | 65.0 | 16.12 | 8.10 | 12.13 | 15.67 | 25.20 | 12.30 | 18.50 | 20.13 | 15.25 | 10.13 |
| <C> | 0.912 | 0.813 | 0.912 | 0.812 | 0.910 | 0.876 | 0.950 | 0.876 | 0.910 | 0.923 | 0.987 | 0.887 | 0.988 | 0.865 |

TABLE II. In each forum, number of communities per teorethical (T), practical excercises (P) and News and Questions Forums. [8]

| | *F1* | *F2* | *F3* | *F4* | *F5* | *F6* | *F7* | *F8* | *F9* | *F10* | *F11* | *F12* | *F13* | *F14* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | 2 | - | - | 2 | - | - | 3 | 3 | 2 | 2 | - | - | - | - |
| P | - | 5 | 4 | - | 5 | 4 | - | - | - | - | 6 | - | - | - |
| NQ | - | - | - | - | - | - | - | - | - | - | - | 2 | 2 | 2 |

[8] Mouronte-López, M. L. (2020). Poster: Studying The Social Networks in Educational Forums. In ICCGI 2020. The Fifteenth International Multi-Conference on Computing in the Global Information Technology

# CONCLUSIONS

- It has also identified the more participatory persons as well as the position that each of them occupies in the network as a whole (power relationships), which has been carried out through the analysis of different types of centrality (Betweenes, PageRank, Degree, EigenVector, Degree).
- Several groups of persons which are especially cohesive have also been detected. These persons and groups had a decisive influence on the results, particularly in the practical exercises.
- The forums related to news and general questions as well as those which refer to theorical contents presented less participation and communities.
- All forums were characterized by a low minimum distance between nodes, which facilitated the propagation of the answers and solutions. High average degree and assortativity between nodes existed. The results allow carrying out improvements in the educational contents and the students’ assessment (participation and involment).

IARIA

Thank you very much for your attention!

Mary Luz Mouronte López

maryluz.mouronte@ufv.es

Higher Polytechnical School

Universidad Francisco de Vitoria

The Fifteenth International Multi-Conference on Computing in the Global Information Technology
ICCGI 2020
October 18, 2020 to October 22, 2020 - Porto, Portugal

# STUDYING THE SOCIAL NETWORKS IN EDUCATIONAL FORUMS

Universidad Francisco de Vitoria
UFV Madrid

**Mary Luz Mouronte-López[1]**

*Higher Polytechnical School at Universidad Francisco de Vitoria*
*Carretera Pozuelo a Majadahonda, Km 1.800 Módulo 1, Carretera Pozuelo a Majadahonda, Km 1.800, 28223 Madrid*

This research aims to carry out a topological study of social networks [4] located in university forums of the Moodle platform. The graphs of the several forums of subjects were built visualizing the structure of the nodes and links and calculating statistical parameters such as: degree, betweenness centrality, clustering coefficient, PageRank, EigenVector centrality [1] and assortativity [2]. The communities' structure was also estimated. This study analyzed how students and faculties worked and socialized in the educational environment, which helped to know more precisely the level of involvement of each student as well as to improve some learning and methodological aspects. Several subjects and forums were analyzed (theoretical and practical contents) . A large amount of data had to be processed.

## CHARACTERISTICS OF THE EDUCATIONAL FORUMS

14 forums each with an average of 115 students were studied. Three types of forums exist: News and questions forums, practical exercise forums and theoretical content forums.

## ANALYZING THE SOCIAL NETWORKS

The XML file of the Moodle forums was analyzed and processed using software programs implemented in Python [5]. In particular, the package networkx was used. The interactions in each forum were represented in a graph *G = (E; S)*, where *E* is the set of nodes corresponding to students and faculties and *S* is the set of links between them. In the following parameters were calculated:

$$b(n) = \sum_{u \neq n \neq w} \frac{\sigma_{uw}(n)}{\sigma_{uw}}$$

Where
$\sigma_{uw}$ is the total number of shortest paths from node $u$ to node $w$
$\sigma_{uw}(n)$ is the number of those paths that pass through n

Clustering coeffient C(n) of a node n Є G

$$C(n) = \frac{2 * t(n)}{d(n) * (d(n) - 1)}$$

Where
$t(n)$ is the number of triangles containing n.
$d(n)$ is the degree of n

Assortativity of a network evaluates the probability of connection between pairs of nodes [2].

EigenVector centrality of a node n Є G :

$$x_n = \frac{1}{\lambda} \sum_{j=1}^{N} x_j = \frac{1}{\lambda} \sum_{j=1}^{N} Aij * xj$$

Where
$Aij$ is element $ij$ of the Adjacency Matrix, such as $Aij$= 1 if node $i$ is attached to node $j$ and 0 otherwise.
This equivalent to $A * X = \lambda * X$ where $\lambda$ is the largest eigen value associated with $A$ and $X$ is its associated eigenvector.

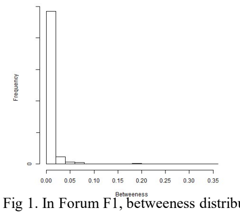
PageRank of a node n Є G :

$$PR(n) = (1 - \alpha) + \alpha * \sum_{w \in V : w \to n} \frac{PR(w)}{k_{out}(w)}$$

Where:

- $\alpha$, damping parameter, $\in$ [0,1]
- $PR(w)$ is the PageRank of the node $w$ which is linked to $n$.

- Degree, Betweeness, Clustering, EigenVector centrality and PageRank distributions



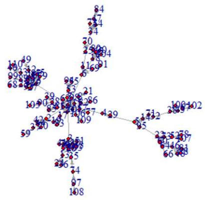



Fig 1. In Forum F1, betweeness distribution

| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 | F11 | F12 | F13 | F14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| <l> | 1,22 | 1,13 | 1,34 | 1,01 | 1,78 | 1,15 | 1,85 | 1,11 | 1,13 | 1,18 | 1,02 | 1,15 | 1,85 | 1,94 |
| <b> | 0.007 | 0.007 | 0.012 | 0.009 | 0.007 | 0.013 | 0.008 | 0.006 | 0.015 | 0.013 | 0.006 | 0.008 | 0.010 | 0.002 |
| <PR> | 0.0002 | 0.0005 | 0.0090 | 0.0031 | 0.0042 | 0.0096 | 0.0063 | 0.0036 | 0.0107 | 0.0114 | 0.0043 | 0.0072 | 0.0078 | 0.0017 |
| <EV> | 0.0018 | 0.0013 | 0.0017 | 0.0078 | 0.0067 | 0.0100 | 0.0088 | 0.0056 | 0.0013 | 0.0238 | 0.0054 | 0.0086 | 0.0095 | 0.0025 |
| <K> | 16,10 | 15,01 | 13,01 | 65,0 | 16,12 | 8,10 | 12,13 | 15,67 | 25,20 | 12,30 | 18,50 | 20,13 | 15,25 | 10,13 |
| <C> | 0,912 | 0,813 | 0,912 | 0,812 | 0,910 | 0,876 | 0,950 | 0,876 | 0,910 | 0,923 | 0,987 | 0,887 | 0,988 | 0,865 |

In each forum, averagre mínimum distance between nodes <l>, average betweeness <b>, average PageRank <PR>, Average EigenVector centrality <EV>, Average Degree <K> and Average Clustering <C> values.

## COMMUNITIES

We also measure the similarities between vertices by means of *Walktrap Algorithm* [3] which uses random walks on *G* to identify communities. This method creates a sequence of partitions $(\mu_k)_{1 \leq k \leq n}$, and chooses the best partition of the network, calculating $Q_k$ for each partition and selecting the partition that maximizes this parameter. The modularity *Q* is defined as the fraction of edges within communities minus the expected value of the same quantity for a random network..

| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 | F11 | F12 | F13 | F14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | 2 | - | - | 2 | - | - | 3 | 3 | 2 | 2 | - | - | - | - |
| P | - | 5 | 4 | - | 5 | 4 | - | - | - | - | 6 | - | - | - |
| NQ | | | | | | | | | | | | 2 | 2 | 2 |

In each forum, number of communities per teorethical (T), practical excercises (P) and News and Questions Forums

## CONCLUSIONS

The research allows to establish a methodology to analyze the interactions between students and faculties in educational forums. The density and cohesion of the components have been studied. It has also identified the more participatory persons as well as the position that each of them occupies in the network as a whole (power relationships), which has been carried out through the analysis of different types of centrality (Betweenes, PageRank, Degree, EigenVector, Degree). Several groups of persons which are especially cohesive have also been detected. These persons and groups had a decisive influence on the results, particularly in the practical exercises. The forums related to news and general questions as well as those which refer to theorical contents presented less participation and communities. All forums were characterized by a low minimum distance between nodes, which facilitated the propagation of the answers and solutions. High average degree and assortativity between nodes existed. The results allow carrying out improvements in the educational contents and the students' assessment (participation and involvment).

## REFERENCES

[1] Newman, M. E.J. (2003). "The structure and function of complex networks", SIAM Review, 45, 167-256.
[2] Newman, M.E.J. (2002). "Assortative Mixing in Networks". Physical review letters, 89 (20), 48109–1120.
[3] Pons, P. And Latapy, M. (2006). Computing communities in large networks using randomwalks. Journal of Graph Algorithms and Applications, 10 (2), 191–218.
[4] Borgatti, S.P., Everett, M.G. and Johnson, J.C. (2018). Analyzing Social Networks, California, U.S.A. : SAGE Publishing.
[5] Python (2020). Python Lenguaje de programación. https://www.python.org/. Accessed July 1, 2020.