Employee effort and productivity in organizations: A network data science approach

**Nan Wang; Evangelos Katsamakas**

# Abstract

Estimating the performance of employees is an important consideration in all organizations. This paper proposes a network data science approach to the estimation and visualization of employee effort, productivity and collaboration patterns. Using data from a software development organization, a bipartite weighted network is firstly constructed for developers’ commitment to project repositories. It is afterwards projected into two weighted one-mode networks: developer-to-developer and repository-to-repositories. For the former graph, two developers are defined to be connected if they once shared projects and for the latter, two repositories relate with each other if sharing developers. Techniques applied include graph theoretic metrics, power-law estimation, and community detection algorithms. We validate the existence of power-law relationships on project sizes (number of developers). We discuss implications for managers and future research directions. As a methodological contribution, the paper demonstrates how network data science can be used to derive a broad spectrum of insights about employee effort in organizations.

**Keywords:** networks, data science, network analysis, power laws, employee effort, productivity, collaboration

# Introduction

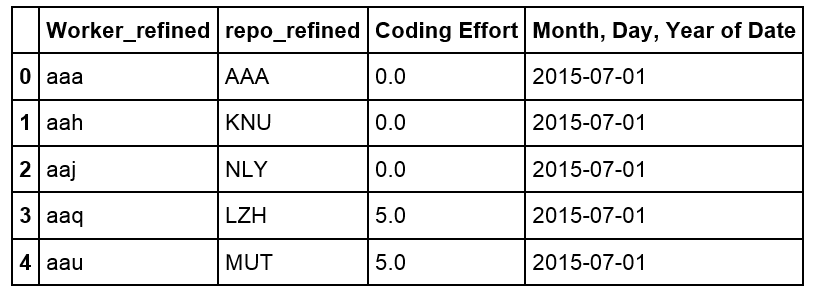
Software application development has turned into an enormously profitable business, with revenue from mobile application purely expected to exceed fifty billion USD by 2016 [1]. Accurate evaluation of developers’ achievement and contribution has been proved to be critical for the long-term development of companies as it is directly related to employee morale, overall productivity and creativity. Poor capability to discover and monitor coding performance prevents technical companies from transforming information on network activity and infrastructural capabilities into strategic knowledge [2, 3].

Although extensive historical study has applied social network analysis as a powerful solution to help understand organizational network performance and service interaction [4], most of them were designed in the context of binary networks. However, many real-life networks are widely recognized to be intrinsically weighted. Characterizing features and structures in weighted networks, therefore, has more important practical significance [5]. Bearing that in mind, this paper investigates a weighted bipartite network where developer is one mode and repository is another.

Question attempted to be answered, at either individual or corporate levels, include identification of developers’ role, contribution, motivation, functionality and working preference, as well as in a broader sense, project membership, corporate cohesion and software development.

# Dataset and data preparation

The dataset used in this paper is the coding effort of 2621 developers to 1705 repositories in a real-life company in a time period of 92 days, from July 1, 2015 to September 30, 2015. This panel data has four variables, namely date, Worker ID, coding effort, Repository ID. There are 172,354 records in total, where “Worker ID” and “Repository ID” identify each unique developer and project respectively. Data is gathered on past software development activity via source code repositories like Subversion and Git, and task tracking systems such as Jira. This dataset is kindly provided by the company BlueOptima. Figure 1 shows an extract from the dataset.



**Figure 1.** Sample of original dataset. Coding effort is a proprietary software measurement and is calculated through evaluating every change that software developers contribute to projects in terms of a series of metrics, such as volume, complexity and interrelatedness of codes [6].

**~~Figure 2.~~** ~~Sample of data (after preparation/aggregation) used in our network data modeling and analysis that follows.~~

# Modeling Contribution: Developers-projects graph

When modelling relations between two different classes of objects, bipartite graphs arise naturally [7, 8]. A graph of customers and products, with an edge between a customer and a product, if the customer has bought that product, is a simple example of a bipartite graph.

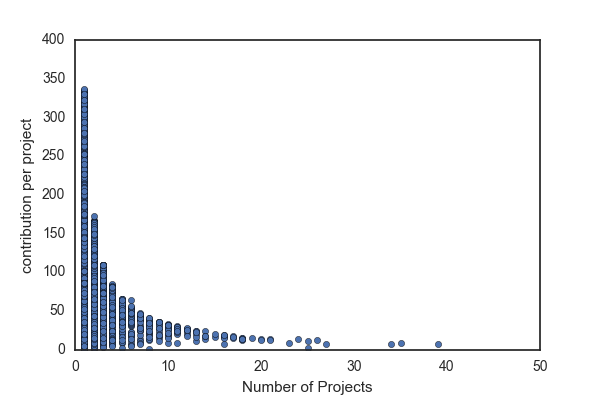
The first network modeled in this paper is a bipartite graph (developer-to-repository network) where developers belong to one mode and project repositories another. Edges, in this case, represent coding effort.

* 1. **Graph Properties**

To understand the properties of this network, we applied various measurements, namely degree centrality, betweeness centrality, closeness centrality and eigenvector centrality [9,10] to the graph. Degree is the number of ties that a node has; Closeness is the average length of the [shortest path](https://en.wikipedia.org/wiki/Shortest_path_problem) between the node and all other nodes in the graph; Betweenness equals to the number of shortest paths from all vertices to all others that pass through that nodes; Eigenvector assumes that each node's centrality is the sum of the centrality values of the nodes that it is connected to [11,12,13].

In the developer-to-repository graph, nodes with degree of one can be interpreted as “focused Exclusive developers” who worked on only one project, or individual project which contained only one developer. It is found that during the entire period in concern, over 50% developers (1409 out of 2459) devoted to only one project and above 40% projects (610 out of 1496) were individual projects. In addition, the maximum degree is 39 for all developer nodes and 96 for repository nodes, meaning that the most multiple-tasking developer (Developer “ouj”) contributed to 39 projects and the largest project (Project “KGD”) were contributed by 96 developers. Also, the average degree is 2.3 for worker nodes and 3.8 for repository, meaning that, overall, each worker worked on two projects and each repository contained four workers.

Therefore, it is reasonable to conclude that sizes of projects, with most being small and a few large, were polarized, implying that the company’s software development business was a combination of a few principal projects and plenty of small projects. Furthermore, if developers’ roles can be implied by the number of projects they worked on and the contribution they made, for example, managers tend to distribute their contribution in a wider range of projects, we are probably able to see the workforce allocation. implications about developers’ role can be made by combining nodes degree and edge weight. As it shown in Figure 3, developer nodes with large degree (many projects) but small-weighted edges (small contribution per project) can possibly be interpreted as project managers, while nodes with small degree (small project volume) and small-weighted edges (small contribution per project) can be considered as unproductive.



**Senior manager**

**Manager**

**Unproductive**

**Figure 3.** Roles and productivity based on node degree and edge weight

* 1. **Graph-Based Recommendation System**

**[https://www.kernix.com/blog/an-efficient-recommender-system-based-on-graph-database\_p9]**

Recommendation systems have been created for various data science applications in a variety of areas. For example, Facebook utilize recommendation systems to suggest friends to users, iTunes and YouTube utilize similar machine learning and recommendation algorithms to suggest songs, videos and movies. Given this general theme, we attempted to create a recommendation system to optimize the assignment of developers to projects.

Collaborative filtering, one of the most popular ways to implement recommendation system, is modeled on the massive users’ preferences and tastes to predict a user’s interest. The underlying assumption of this approach is that if a person *A* has the same opinion as a person *B* on an issue, A is more likely to have B's opinion on a different issue than that of a randomly chosen person. There are three types of collaborative filtering algorithms in the recommendation system literature, namely Memory Based Approaches, which finds similar users or items based on historical behaviors or records, Model Based Approaches, which makes prediction from probabilistic models or some sort of machine learning algorithms, as well as Hybrid Approaches, which combines the memory based and model based approaches to take the advantages of both approaches. Each of those approaches has its own advantages and limitations.

Considering that our data only contains 92 days’ records and provides no worker preference information, we turned to a graph-based approach which searches similar users or items as ‘neighbors’ on the graph. In the developer-project case, two developers are neighbors if they share projects. To define and calculate the similarity between neighbors, the following function was applied. The similarity between two developers u1 and ui can be expressed as:

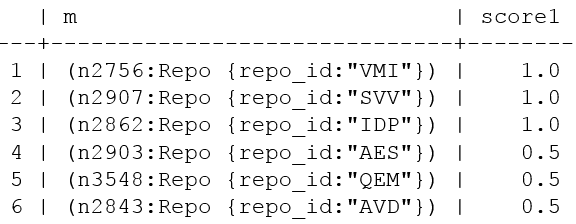
Number of Common Projects between u1 and ui

Sim(u1, ui) =

Number of Projects of ui

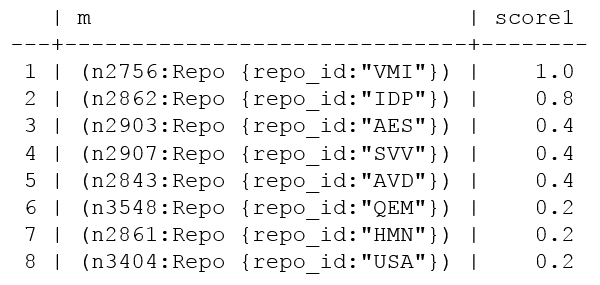
Therefore, the similarity between two developers is the ratio of the number of projects they have in common over the number of projects they work on. We further set a threshold of 0.5 for similarity, which means that developer A is considered to be similar to developer B only if B work on at least half of A’s projects.

Below is a list of projects recommended by the system to the developer ‘xev’:

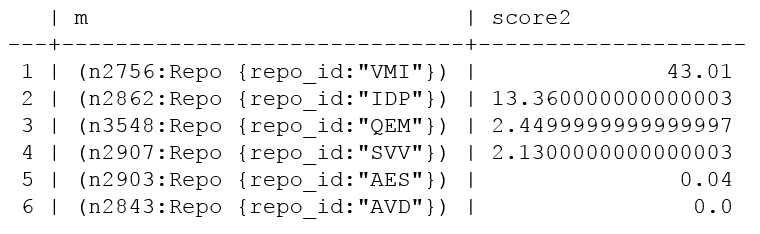


The score is the ratio of similar neighbors who contribute to the recommended project over those who don’t. Therefore, new projects are recommended to the developer ‘xev’ in an order from ‘worked by most similar neighbors to ‘worked by fewest similar neighbors. Accordingly, assuming a manager is about to assign the developer ‘xev’ a project suitable for his skills, project ‘VMI’, ‘SVV’ and ‘IDP’ would be worthy of consideration, given the fact that all the developer’s similar neighbors work on them.

It is worth mention that different similarity threshold would affect recommendation results largely. For example, if we change the similarity threshold to 0.4, the results would be as follows and ‘VMI’ is the only project shared by all the developer’s acquaintances.



Below is another list of projects recommended by the system to the developer ‘xev’ with the similarity threshold being 0.5. Different from the previous system, the scoring function has changed into the sum of similar neighbors’ contribution. Therefore, the system recommends new projects to the developer ‘xev’ in an order from ‘contributed the hardest by similar neighbors’ to ‘contributed the least by similar neighbors’. If a manager wants to ensure developers’ contribution when assigning new projects, he probably prefers this recommendation scheme based on the assumption that a developer will contribute a lot to a project if similar developers unanimously work hard on that project.



An optimized graph-based recommendation system is beyond scope of this paper. Here we just implemented a general-purpose way to demonstrate the efficiency and simplicity of a graph-based approach. Unlike model-based approaches, graph-based approaches doesn’t necessarily require user features or large volume of data for training. It has been proved to be able to solve the problem of data sparsity. [https://arxiv.org/ftp/arxiv/papers/1604/1604.03147.pdf]

* 1. Conclusion

With the weighted bipartite developer-project graph, we had access to gain an overview of the company’s software development business and workforce allocation, as well as build a graph-based recommendation system which helps project assignments with different purposes.

# Modeling Collaboration: Developers-developers graph

While limited methods exist for analyzing two-mode networks, transforming a two-mode network into a one-mode network, often referred to as projection, is necessary [16]. The process works by selecting one set of nodes, and linking two nodes if they are connected to the same node of the other set. However, since one-mode projection is always less informative than the bipartite representation and the redistribution of weights would strongly affect the community structure [17], a proper weighting method is required to better retain the original information.

In this paper, Newman’s weighted projection was applied, which adopts the weighting scheme presented by Newman~~. The formula is as follows [18]:~~ This method would create a one-mode network in which the out-strength of a node is equal to the sum of the weights attached to the ties in the two-mode network that originated from that node.

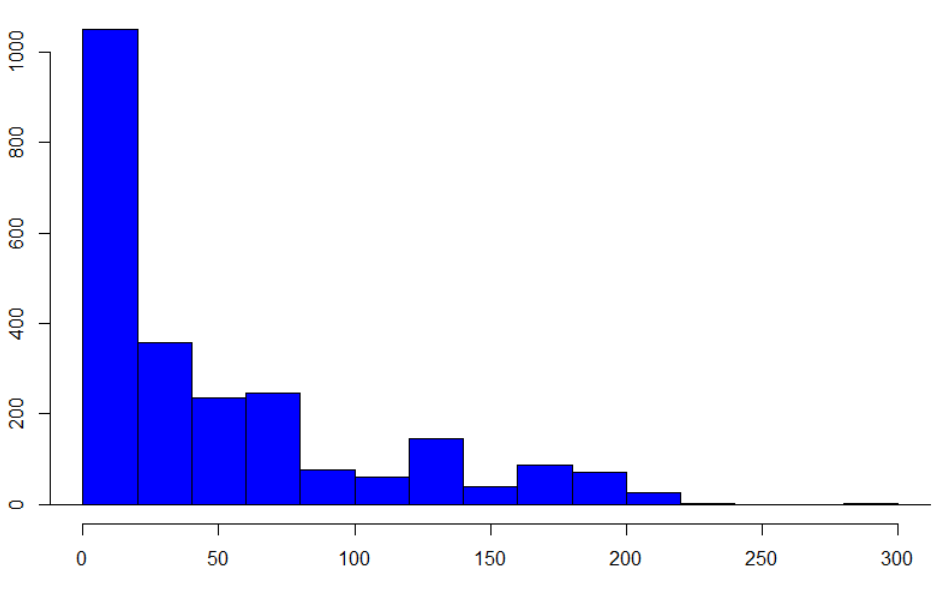
Accordingly, the developer-to-repository graph is projected into two separate one mode graphs: a developer-to-developer graph with 2459 nodes and 61378 edges where each developer is a node and an edge exists between two nodes if both developers are on the same project. Another is a repository-to-repository graph with 1496 nodes and 16764 edges where repositories are nodes and edges represent shared developers. ~~Therefore, three weighted networks have been built for analysis in this paper.~~

* 1. **Graph Properties**

We can then explore developers and their connection through studying properties of the developer-developer graph.

* + 1. **Degree Centrality**

We firstly applied degree centrality and plotted the distribution of degrees. Figure 7 shows that the degree distribution is likely to follow power-law distributions. In network theory, a network is named scale-free if its degree distribution follows a mathematical function called a power law [22]. Unlike a random graph (i.e., new nodes attach to existing nodes with uniform probabilities), a scale-free graph displays preferential attachment of new nodes (i.e., some nodes have higher probability of attachment than others) [25]. In our developer-project case, it is intuitive that the network is not randomly formed. It is also in accordance with expectation that a developer is more likely to collaborate with some developers because of shared skillsets or diversified experience levels or project requirements.



**DEVELOPER NETWORK**

**Histogram for node degree**

**Figure 7.** Node degree distribution for developer graph

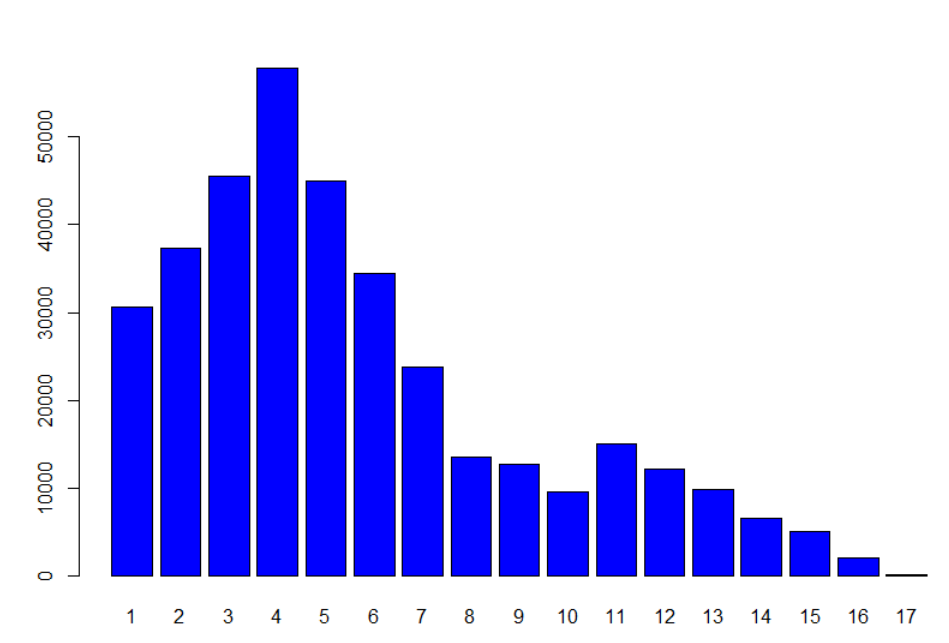
To validate our assumption that nodes in the graph have preferential attachment, we applied a Statistical analysis called Kolmogorov-Smirnov test. [Hazewinkel, Michiel, ed. (2001), ["Kolmogorov–Smirnov test"](https://www.encyclopediaofmath.org/index.php?title=p/k055740), [Encyclopedia of Mathematics](https://en.wikipedia.org/wiki/Encyclopedia_of_Mathematics), Springer, [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [978-1-55608-010-4](https://en.wikipedia.org/wiki/Special:BookSources/978-1-55608-010-4)] Test results are as follows. The combination of small test statistic (KS.stat) and bigger p-value (KS.p) together indicates a good fit of power-law distribution [23, 24]. As it shown in Table 1, KS.stat is 0.074 and KS.p is 0.99. It therefore proves that the developer-to-developer network has the property of scale-free

|  |  |  |
| --- | --- | --- |
| **Measurement** | **Definition** | **Value** |
| Alpha | The exponent of the fitted power-law distribution. | 18.67 |
| xmin | The minimum value from which the power-law distribution was fitted. | 206 |
| logLik | The log-likelihood of the fitted parameters. | -45.63 |
| KS.stat | The test statistic of a Kolmogorov-Smirnov test that compares the fitted distribution with the input vector. **Smaller scores denote better fit.** | 0.074 |
| KS.p | The p-value of the Kolmogorov-Smirnov test. Small p-values (less than 0.05) indicate that the test rejected the hypothesis that the original data could have been drawn from the fitted power-law distribution. | 0.99 |

**Table 1.** Statistical tests for power-laws

* + 1. **Average Path Length**

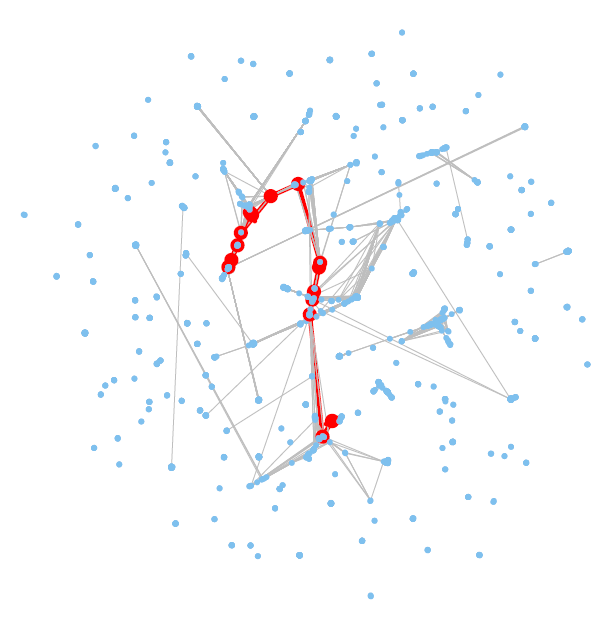
Small world hypothesis states that two arbitrary people are connected by only six degrees of separation [26]. shortest path length is the number of steps along the shortest paths for all possible pairs of network [nodes](https://en.wikipedia.org/wiki/Node_(networking)). Figure 8 shows the distribution of shortest path length in the graph. Averaged at 5.63, lengths of the shortest path from one to seventeen, meaning that a developer may expect to reach a randomly-selected developer in the distance of at least one and at most seventeen in the company of 2,459 developers.



**DEVELOPER NETWOKR**

**Histogram for Shortest Path**

**Figure 8.** Distribution of Shortest Paths in the developers’ network



**Diameter: 17**

**DEVELOPER NETWORK**

**Figure 9.** Visualization of developer network diameter

* + 1. **Clustering Coefficient**

Ranging from zero to one, clustering coefficient is a measurement of nodes clustering tendency of clustering, with zero meaning no clustering and one maximal clustering. There are two types of clustering coefficient: local and global. The average local clustering coefficient places more weight on low degree nodes while the global emphasizes on high degree nodes [29].

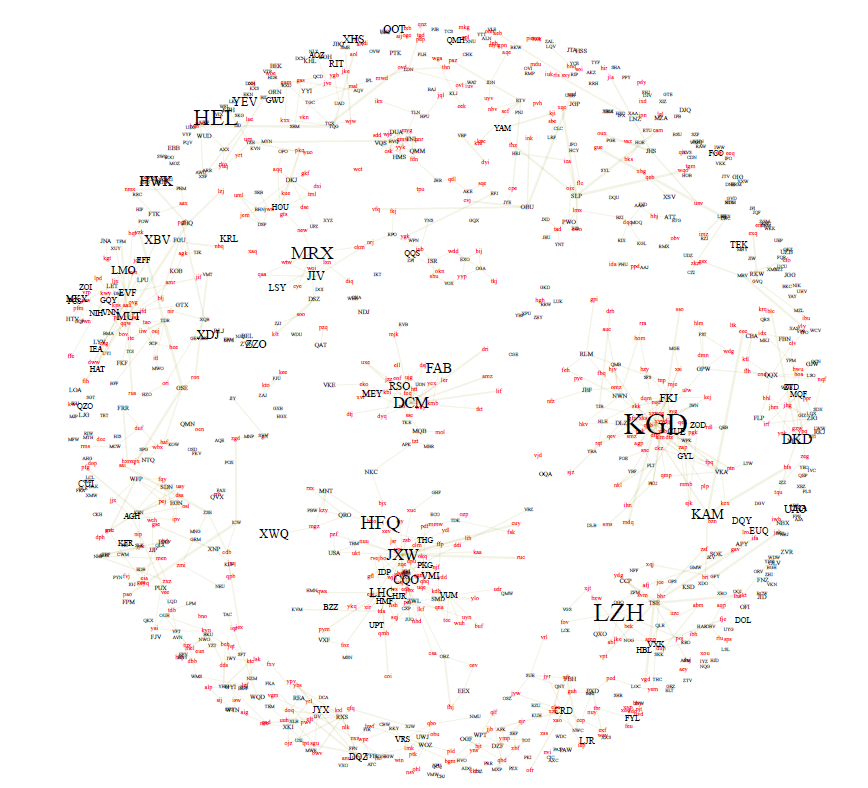
The global clustering coefficient of the developer-developer graph is as high as 0.92, indicating that the network is supposed to form cliques (fully connected clusters) and stay well-connected. However, the average local clustering coefficient turns out to be only 0.21. The largest cluster only consists of 610 developers, 25% of total number of developers. The second largest cluster contains 535, followed by a third of 52. Moreover, 34 out of 142 clusters are composed of only two nodes.

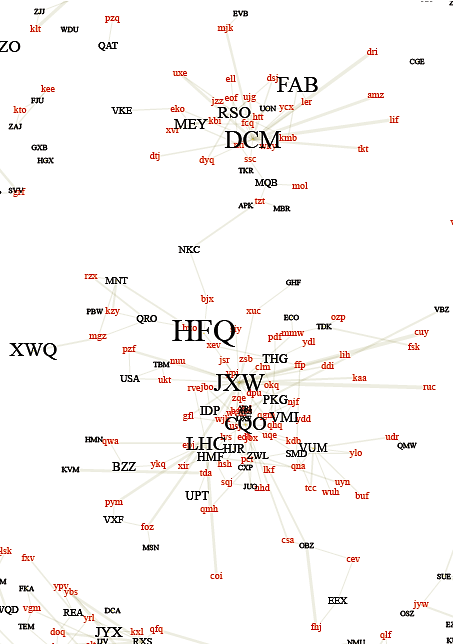
With a large global clustering coefficient and a small local one, we can suppose that bridges facilitating inter-group communication are lost. This could be explained by the company information separation policy or the diversity of developers’ functionality and project purpose. For example, developer using Ruby is less likely to be teamed with developers using Java, and projects for predictive modeling usually share little resource of front-end development.

* + 1. **Betweenness Centrality**

For example, nodes with higher betweenness scores have more control of other nodes in terms of communication access. Some developers have high betweenness centrality but low degree. For example, node of Developer “cmt” has degree of 2 and betweenness of 1596, meaning that although having two connections, Developer “cmt” did play an important role as a go-between for many others.

In Figure 4, “DCM” and “HFQ” were two big projects sharing few developers. In order to achieve information sharing in between, the quickest path would be one from Developer “ssc” to Developer “tzt” and finally to Developer “bjx”. Those three developers, though having a small number of connections, were crucial to any developers in need of information sharing in either project. Therefore, they are nodes have small degree but high betweenness.





**Figure 4.** Visualization of developers (nodes) with high betweeness

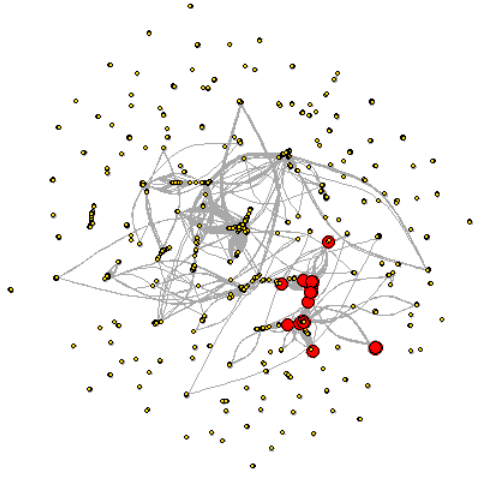
The graph is drawn according to Kamada–Kawai algorithm [15]

Nodes’ label size is set to be positively correlated with nodes’ degree

* + 1. **Closeness Centrality**

Closeness centrality [13] is a measurement of nodes’ access to the resources in the network. High closeness centrality indicates ~~which~~ nodes with the quickest access to most resources in the network. In our case of software development network, developers with high closeness centrality scores therefore play a similar role of gatekeepers ~~in organizational studies~~ for technology diffusion.

As shown in Figure 5, nodes colored in red are those with highest closeness scores and they are densely connected with each other. These developers may need to be identified and then supported or monitored if ~~to~~ controlling the diffusion of ideas and technology between disparate development groups is required.



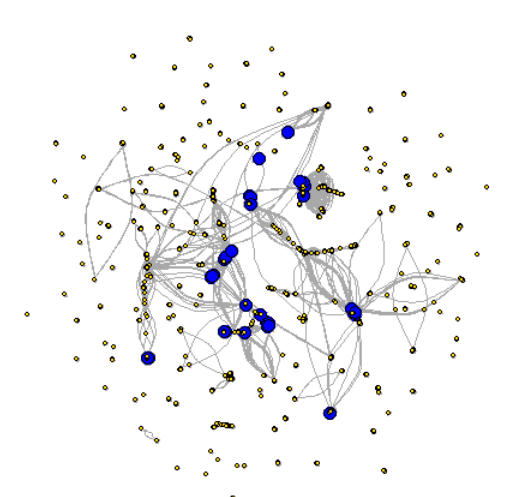
**Figure 5.** Visualization of developers ~~(nodes)~~ with high closeness centrality

* + 1. **Eigenvector Centrality**

Eigenvector Centrality is another measurement decided by both the number of neighbors and the quality of its connections [13]. In other words, a high eigenvector score is resulted from a large number of high-quality neighbors. ~~Surprisingly~~ ~~howeve~~r, In the developer-to-developer network, it was found that 44% of developer nodes with top eigenvector centrality scores has only one degree, meaning that many influential developers only connected with one ~~more~~ person in the network.

As shown in Figure 6, nodes colored in blue are those with highest Eigenvector. Compared to those with high closeness (Figure 5), they ~~were~~ are more widely positioned among different subcomponents in the graph. Based on the observation that the nodes with highest Eigenvector scores are connected with different sets of nodes, we can probably interpret them as managers leading different projects in the network.

~~If we interpret high closeness nodes as social sociable developers who knew many people, nodes with highest Eigenvector scores can be seen as are actually influential developers who might act as supervisors or managers in various departments in the network based on the observation that they are connected with different sets of people.~~



**Figure 6.** Visualization of developers ~~(nodes)~~ with high eigenvector scores

## 4.2 Community Detection

Communities in a network are groups of nodes internally connected or nodes sharing attributes. Detecting communities provides insights regarding overall network structure, behavioral patterns of nodes and their relations. In this case particularly, it would lead to findings about developers’ roles, collaboration preference and their interaction as well as corporate cohesion [31].

This paper chooses *R* package *igraph* [32] as a main tool to conduct community detection. Applicable algorithms in the package include Edge-Betweenness, Leading Eigenvector, Fast-Greedy, Walktrap, Label Propagation and Infomap.Summary and comparison of those algorithms are shown in Table 2.

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Directed Edges** | **Weighted Edges** | **Multiple Components** |
| *Edge-Betweenness* | TRUE | TRUE | TRUE |
| *Leading Eigenvector* | FA LSE | FALSE | TRUE |
| *Fast-Greedy* | FALSE | TRUE | TRUE |
| *Multi-Level* | FALSE | TRUE | TRUE |
| *Walktrap* | FALSE | TRUE | FALSE |
| *Label Propagation* | FALSE | TRUE | FALSE |
| *Infomap* | TRUE | TRUE | FALSE |

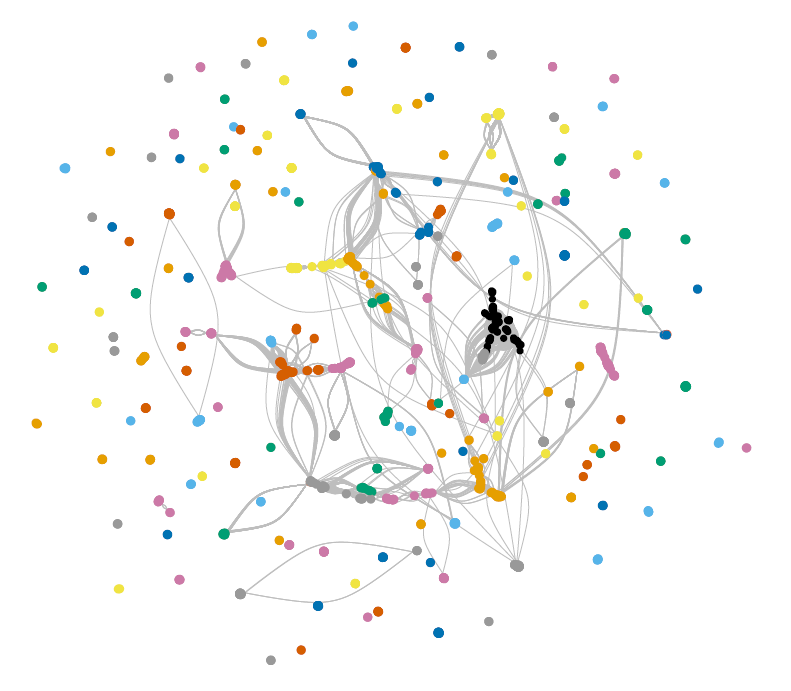
**Table 2.** Community detection algorithms used and their properties

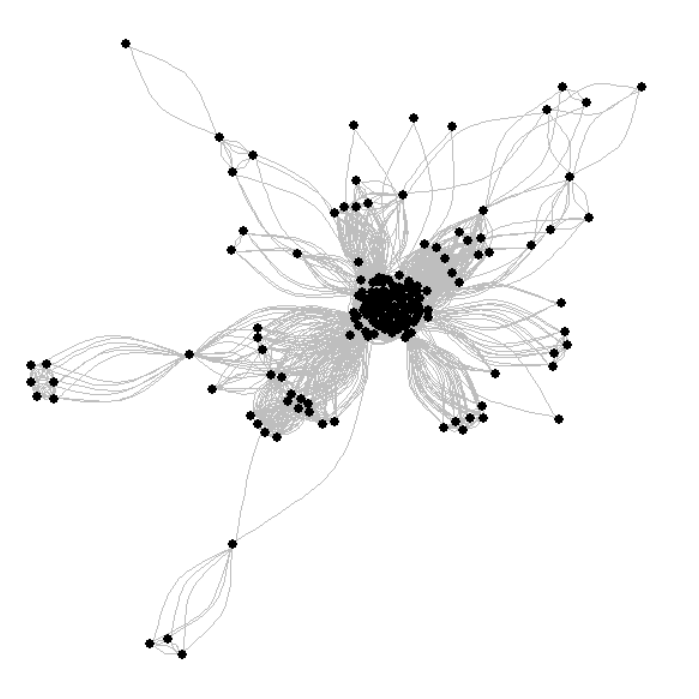
Considering the structured networks are weighted undirected graphs, four out of seven algorithms above have been implemented, namely *Multi-Level*, *Walktrap*, *Label Propagation* and *Infomap*. In order to measure the performance of those algorithms we used *modularity*, a common measure of dividing a network into communities [33]. Performance for each algorithm is shown in Table 3 below. Combining the performance of modularity and group number, it is clear that the *Multi-Level* algorithm works the best, followed by the *Walktrap* algorithm.

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Modularity** | **Groups** |
| *Multi-Level* | 0.97 | 172 |
| *Walktrap* | 0.96 | 193 |
| *Label Propagation* | 0.89 | 288 |
| *Infomap* | 0.16 | 1119 |

**Table 3.** Performance of algorithms used for community detection

According to the *Multi-Level* algorithm, the largest community in this company contain 173 developers while the smallest (six communities) consist of just two developers. On average, communities contain 98 developers. Figure 10 demonstrates 172 colored groups, with the largest group magnified.





**Figure 10.** Visualization of community detection in the developer’s network

# Modeling Dynamics: Evolving Network

Analyzing dynamic network has gained increase popularity and significance as longitudinal network data becomes more available. Organizations study corporate network in a dynamic way in order to discover the decision cycle of major events, track the evolutions of corporate finances, operations and culture, as well as detect changes in the organizational behavior. Being able to identify a change in network dynamics can enable managers to better prepare for the coming change and prevent potential problems. For example, before the introduction of e-mail and cell phones into the workplace, workers had limited peers available that they could ask about problems before they had to seek guidance from senior management. However, with growing on-line communities, the available peers to consult are no longer limited to acquaintances. Individual network is becoming larger. While this is good that workers are able to resolve problems at a lower level, senior managers are unable to influence decisions with their senior

judgment and experience. Dynamic network analysis therefore can provide those managers with a tool to prevent potential problems in their organization by tracking every change in the social network of employees. [http://www.casos.cs.cmu.edu/publications/papers/CMU-ISR-09-118.pdf]

Dynamic network analysis (DNA) brings together traditional social network analysis, link analysis, social simulation and multi-agent systems within network science and network theory. There are two aspects of this field. The first is the statistical analysis of DNA data. The second is the utilization of simulation to address issues of network dynamics. [https://en.wikipedia.org/wiki/Dynamic\_network\_analysis ]. Dominant modeling methods include Markov chain models, multi-agent simulation models, and statistical models. [http://www.casos.cs.cmu.edu/publications/papers/CMU-ISR-09-118.pdf]

A complete application of dynamic network modeling is beyond the scope of this paper. Considering the nature of the given problem, we choose statistical models as our major measure. In other words, we compare the properties of networks at different points in time.

# Discussion and concluding remarks

# A study of the proprietary software development was started with constructing weighted graphs. Afterwards, initial insights about developers’ role (managers or engineers), productivity and individual influence were achieved through analyzing network properties. Finally, implications about software development cohesion and culture were developed through implementation of cluster analysis as well as a comparison between constructed networks and complex systems.

# Several assumptions and limitations are present in the study. For example, the analysis only looked at a snapshot or aggregated version of the software development network at this particular company. A longitudinal and dynamic analysis in the future may provide better understanding of how node attach and detach from the network (developers dropping off or initiating projects), and thus lead to a more thorough understanding of developers’ working pattern.

# References

[1] Statista. (2014). Worldwide mobile app revenues from 2011 to 2017 (in billion U.S. dollars). Retrieved from <http://www.statista.com/statistics/269025/worldwide-mobile-app-revenue-forecast/>

[2] McGarry, John (2001) Practical Software Measurement, p19. Addison-Wesley.

[3] UKESSAYS (19) Importance of Software Measurement and Metrics. Retrieved from <https://www.ukessays.com/essays/information-technology/importance-of-software-measurement-and-metrics-information-technology-essay.php>

[4] Wu, M.W., Lin, Y. D. (2001). "Open Source Development: An Overview," IEEE Computer, pp. 33-38.

[5] S.Z. Yang, S.W. Luo. (2009). Modern Physics Letters B 23, 3209.

[6] Newswire (2013). BlueOptima Coding Effort Analytics Announces Support for Git. Retrieved from <https://www.newswire.com/blueoptima-coding-effort-analytics/252989>.

[7] Asratian, Armen S., Denley, Tristan M. J., Häggkvist, Roland. (1998). Bipartite Graphs and their Applications, Cambridge Tracts in Mathematics, 131, Cambridge University Press, ISBN 9780521593458.

[8] Guillaume, J., Latapy, M. (2004). Bipartite structure of all complex networks. Information Processing Letters, 90(5):215-221.

[9] Newman, Mark, Albert-László Barabási, Watts, Duncan J. (2006). The Structure and Dynamics of Networks, The Princeton Press, ISBN 0-691-11357-2

[10] Newman, M.E.J. (2010). “Networks: An Introduction.” Oxford, UK: Oxford University Press.

[11] Bonacich, Phillip (1987). "Power and Centrality: A Family of Measures". American Journal of Sociology. University of Chicago Press. 92: 1170–1182. doi:10.1086/228631.

[12] Borgatti, Stephen P. (2005). "Centrality and Network Flow". Social Networks. Elsevier. 27: 55–71. doi:10.1016/j.socnet.2004.11.008.

[13] Borgatti, Stephen P.; Everett, Martin G. (2006). "A Graph-Theoretic Perspective on Centrality". Social Networks. Elsevier. 28: 466–484. doi:10.1016/j.socnet.2005.11.005.

[14] Benzi, Michele; Klymko, Christine (2013). "A matrix analysis of different centrality measures". arXiv:1312.6722Freely accessible.

[15] Kamada, Tomihisa, Kawai, Satoru (1989), "An algorithm for drawing general undirected graphs", Information Processing Letters, Elsevier, 31 (1): 7–15, doi:10.1016/0020-0190(89)90102-6

[16] Zhou, Tao, Ren, Jie, Matúš Medo, Zhang, Yi-Cheng. (2007). "Bipartite network projection and personal recommendation". PHYSICAL REVIEW E 76(4): 046115.

[17] Fan, Ying, Li, Menghui, Zhang, Peng, Wu, Jinshan, Di, Zengru. (2007). "The effect of weight on community structure of networks". PHYSICA A 378. 583–590.

[18] Klein, David R. (2012) "Chapter 4 Section 6: Drawing Newman Projections." Organic Chemistry. 2nd ed. N.p. Wiley. 158-59. Print.

[19] Ladyman, James; Lambert, James; Wiesner, Karoline (2013). "What is a Complex System?". European Journal for Philosophy of Science. 3: 33–67. doi:10.1007/s13194-012-0056-8.

[20] Simon, Herbert A. (1991). "The architecture of complexity". In George J. Klir. Facets of Systems Science. Springer US. pp. 457–476. doi:10.1007/978-1-4899-0718-9\_31.

[21] Albert R.; Barabási A.-L. (2002). "Statistical mechanics of complex networks". Rev. Mod. Phys. 74: 47–97. arXiv:cond-mat/0106096Freely accessible. Bibcode:2002RvMP...74...47A. doi:10.1103/RevModPhys.74.47.

[22] Clauset, A.; Shalizi, C. R., Newman, M. E. J. (2009). "Power-Law Distributions in Empirical Data". SIAM Review. 51 (4): 661–703. arXiv:0706.1062Freely accessible. Bibcode:2009 SIAMR..51..661C. doi:10.1137/070710111.

[23] Stephens, M. A. (1974). "EDF Statistics for Goodness of Fit and Some Comparisons". Journal of the American Statistical Association. American Statistical Association. 69 (347): 730–737. doi:10.2307/2286009. JSTOR 2286009.

[24] Marsaglia G, Tsang WW, Wang J (2003). "Evaluating Kolmogorov's Distribution". Journal of Statistical Software. 8 (18): 1–4.

[25] Béla Bollobás, Random Graphs, 2nd Edition, 2001, Cambridge University Press

[26] Barabási, Albert-László (2004). Linked: How Everything is Connected to Everything Else. ISBN 0-452-28439-2.

[27] Holland, P. W., Leinhardt S. (1971). "Transitivity in structural models of small groups". Comparative Group Studies. 2: 107–124.

[28] Opsahl, Tore, Panzarasa, Pietro. (2009). "Clustering in Weighted Networks". Social Networks. 31 (2): 155–163.

[29] Kaiser, Marcus. (2008). "Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks". New Journal of Physics. 10 (8): 083042. arXiv:0802.2512Freely accessible.

[30] Watts, D. J., S. H. Strogatz, (1998). Nature (London) 393,440

[31] Girvan, M., Newman, M. E. J. (2002). "Community structure in social and biological networks". Proc. Natl. Acad. Sci. USA. 99 (12): 7821–7826. doi:10.1073/pnas.122653799. PMC 122977Freely accessible. PMID 12060727

[32] Gábor Csárdi, Tamás Nepusz. (2006). “The igraph software package for complex network research. InterJournal Complex Systems.” necsi.edu.

[33] Baldwin, C. Y., Clark, K. B. (2000). “Design rules, Volume 1: The power of modularity.” Cambridge, MA: MIT Press.