A Network Analysis on Movie Producing Teams and their Success

Wladston Viana, Pedro Onofre Santos, Ana Paula Couto da Silva, Mirella M. Moro
Computer Science Departament
Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
Email: {wladston,pedroo,ana.coutosilva,mirella}@dcc.ufmg.br

***Abstract*—We perform social network analysis on movie producing teams formed by directors, producers and writers, using data from IMDb. We assemble an evolving social network by linking agents that worked together throughout history. After, we proceed to calculate topological and non-topological metrics from this network and its teams through time. We present the evolution of topological and non-topological metrics. We analyze the correlation between these metrics and two success parameters: movie's ratings and gross income.**

***Keywords*-Social Networks, Topological and Non-Topological Metrics, Team Success, IMDb.**

I. Introduction

A wide range of human interactions and relationships can be represented through graphs. Besides the well-known online social networks such as Twitter[1] and Facebook[2], social network analysis techniques are also being applied to study collaboration among actors, athletes, executives, musicians, scientists, and many other work environments involving teamwork and group-oriented activity [1], [5], [8], [9], [13], [14].

It is on the best interest of managers and policy makers to form teams in such way that maximizes the productivity. The director of a soccer club would benefit from composing a team with higher winning odds; a faculty director would want to fund research teams that will produce better articles; a manager from a company wants to re-arrange his team in order to ramp up productivity. To help in these important tasks, a huge amount of data regarding how groups work and interact became available in the latest years. With this data, new research revealing previously unknown correlations between properties from the underlying interaction network and the overall success and output quality from its agents are being proposed.

In the entertainment industry, agents team up and work together in order to produce movies, television shows, music albums, Broadway musicals, and many more. Among all entertainment branches, the context in this work is on the film industry. The movie industry per se is a billion dollar business; hence, a movie's public acclaim and critic review play a very important economic role. Using data from the Internet Movie Database (IMDb[3]) to analyze metrics based on team composition and arrangement in the network, we might discover factors associated with the production of better movies. The IMDb is one of the most thorough and detailed cinema database over the internet. An analysis of such extensive data yields more robust and reliable conclusions than many previously conducted experiments performed over smaller data [14].

[1]Twitter: http://www.twitter.com
[2]Facebook: http://www.facebook.com
[3]IMDb: http://www.imdb.com

Note that accessing the relation between topological and non-topological properties of a collaborative network and its success parameters has a high relevance for any industry. Specifically, such relation may guide strategies for organizing teams in a way that optimizes their revenue capacity and social impact. In this work, we study known topological metrics (such as the small world coefficient, betweenness, closeness and local clustering coefficient) applied over the IMDb data for the movie industry. Some metrics are global and relative to the network as a whole, whereas others are local and specific to agents in a single movie producing team. We also study some non-topological metrics, such as past individual experience. We then correlate these metrics with movie's success parameters (rating and gross income).

Next, we discuss the related work (Section II) and the dataset that we analyze (Section III). Then, we go over our main contributions, which are summarized as follows:

- We describe our network model for movie-producing teams composed by producers, directors and writers. We also define topological and non-topological metrics for studying the impact of team composition in the movie success (Section IV).
- We experimentally analyze the correlation between topological and non-topological metrics with movies' rating and gross income success (Section V).

II. Related Work

Understanding how people work together in order to better achieve goals has been explored in many different contexts [2], [3], [4], [5], [6], [8], [9], [11], [12], [13], [14]. Many research papers focus on team formation among scientists and their publication rate and impact factor metrics. For instance, scientific collaboration networks and their properties have been studied by Newman [8], [9]. The author shows that different scientific communities form small-world networks and are highly clustered, and proposes a method for estimating tie strength. Borner et al.[13] explores the

978-1-4799-6953-1/14 $31.00 © 2014 IEEE
DOI 10.1109/LAWeb.2014.10

"Science of team science", a research area focusing on the processes by which scientific teams organize and conduct their work. Such research explores how teams connect and collaborate in order to achieve breakthroughs that would not be attainable by either individual or simple additive efforts.

People also aspire to understand factors that may explain high productivity and success across many scenarios, making network studies of team formation go beyond collaboration among scientists. For instance, Nemoto et al.[6] showed that Wikipedia[4] editors with more social capital (taking part in a cohesive and centralized cluster) produce higher quality articles faster. Singh et al.[12] found that specific kinds of network ties among open source developers are correlated with the development of more popular open source projects.

Other authors explore the network topology of the agents as a tool for understanding their success. Most of them study the correlation between success and the small world coefficient of the network. Chen et al.[2] studied the network formed by collaboration among countries and showed that the small world coefficient is correlated to patent registrations. Schilling and Phelps [11] studied the collaboration among companies and found that the small world metric is correlated to knowledge creation inside companies.

Regarding the entertainment segment, the work by Uzzi and Spiro [14] is the most related to ours. The authors studied the network formed by Broadway musical producers (choreographers, writers and directors, not the cast), and found evidence that the artistic and financial success of such a network as a whole is correlated to its small world coefficient. The authors analyzed many network metrics and found that some of those were correlated to success, while others were not.

To the best of our knowledge, the present work is the first to study the relation between network aspects and success considering motion pictures producers. Furthermore, the dataset is large and composed by several movie genres.

III. Dataset Description

In this work, we analyze the IMDb database, which contains information from thousands of movies from the late 1800's until 2013, from all over the world. For each movie, its list of directors, writers and producers is available, as well as the rating received from IMDb's users. For some movies, the gross income is also available. It is important to state that only movies produced for cinema were analyzed, leaving all TV productions out of the experiment: TV productions are essentially organized differently than cinema productions, and it is debatable if ratings from TV series and movies can be compared to cinema ratings. In total, over 190 thousand of cinema titles were available from the database at the time it was fetched, containing over 320 thousand production team members (directors, writers and producers).

[4]Wikipedia: http://www.wikipedia.org

Most of the movies in IMDb are from extremely unknown productions, which received very little or no user ratings and reviews. Ratings for those movies cannot be compared to well established cinema productions, therefore we decided to filter out those that received less than 25 thousand user votes. That is also a prerequisite for inclusion in the *IMDb TOP250 list* [5], and it clearly selects movies with substantial social impact. We compared this subset of the database with the whole, and it still maintains a similar histogram of number of productions per year, user votes per year and number of agents per team. Also, the non-significant movies only add noise to the correlation analysis, i.e., the dataset without them provide more homogeneous sample. The final subset contains about 1.5% movies (3006 titles) of the total [6].

Evaluating metrics on a network with very few nodes and edges may produce distorted results. It is then necessary to bootstrap the movie producing graph until it reaches a minimum size, i.e., before network metrics become significant. For this reason, we use all movie data from before 1945 just to bootstrap the network with edges and vertices. The experimental analysis considers the whole historical network, but the network metrics and movie success parameters were only extracted for movies produced after 1945.

For evaluating the movie's economic success, we chose the gross income, as it is directly connected with the title's financial revenue and represents how many people were interested in paying to watch such a movie. Also, as a public's acceptance metric, we considered the IMDb user rating, as it indicates how well the title was received by the public. Using these two variables, we are also able to correlate the movie's economic success with its public acceptance.

Ratings for the movies were normalized for the number of votes received using a true Bayesian estimate, which is the same used by IMDb in its TOP 250 movie list:

$$WeightedRating = (\frac{v}{v+m}) \times R + (\frac{m}{v+m}) \times C, \quad (1)$$

where, for each movie, R is the mean of its ratings, v is its number of votes received, m is the least possible amount of votes (25 thousand), and C is the mean vote across the whole report. The value of C is provided by IMDb and it is equal to 7.0. For the TOP 250, only votes from regular voters are considered.

The gross income information used in our work is also present in the IMDb database, but only for a few movies. The gross income value is usually given in the currency of the country that hosted the movie production, and is dated from shortly after the movie's release. In order to accurately compare gross income from different movies with minimal

[5]IMDB TOP250 list: http://www.imdb.com/chart/top

[6]We have also evaluated the results considering the (significantly larger) set of movies (more than thousand ratings), comprising about 9% of the total, and the results were similar.

distortion, the values had to be normalized. Monetary figures for gross income were converted to US Dollars using the Historical Currency Converter Web Service[7]. The corresponding amount in US Dollars was subsequently corrected for inflation considering the present time using the CPI Inflation Calculator[8], an online feature provided by the Bureau of Labor Statistics. Gross income figures not listed in US Dollars that also did not possess a valid historical exchange record, were discarded, as they represented only about 0.2% (6 movies) of the chosen sample.

IV. Network Modeling and Topological Metrics

The IMDb database provides the full cast and crew from movies, including actors, producers, directors, writers, art direction, special effects team, soundtracks and sound effects department, and many more. For modeling a movie-producing team, we decided to include only the producers, directors and writers, leaving out the rest of the production crew and cast. This choice was made because such selected agents are the core of the team: they take the important decisions and hire the rest of the crew. The responsibility for the success of the movie ultimately falls on those agents. The total of agents in our network is $11,832$.

We model the IMDb movie database as a bipartite graph, with edges between a set a movies and a set of selected agents (producers, writers and directors), indicating individuals who produced each movie. Most network metrics in the literature cannot be applied to bipartite networks, so in order to calculate them we projected the network into a one-mode graph. In this projection only agents are present as nodes, and edges exist between agents who worked on a same movie, following the methodology proposed by Newman [8].

Since we are interested in studying the network's evolution through time, we process the dataset in chronological order of movie production. For each movie, we take its producers, writers and directors as vertices, and create unweighted edges between them to indicate existing previous work.

To increase the fidelity of our model to how movie producing parties actually interact, when a node ceases to participate in any movie for more than 7 years, we remove it and all its vertices from the database. We note that such an agent is likely to be retired and thus not participating actively in the network, following the same methodology proposed in [14].

In our analysis, we consider the small world coefficient for measuring the overall cohesion in the entire network. The small world coefficient is calculated from two other global network metrics:

(1) Network Clustering Coefficient: The clustering coefficient is the average fraction of pairs of an agent's collaborators who have also collaborated with one another. Mathematically [7]:

$$\mathcal{CC} = \frac{\textit{3} \times \textit{number of triangles in the network}}{\textit{number of connected triples of vertices}}.$$

Here a "triangle" is a trio of agents (producers, writers and directors), each of whom is connected to both of the others, and a "connected triple" is a single agent connected to two others.

(2) Network Average Shortest Path: Let P(a,b) be the set of paths between a given pair of agents a and b. We define the shortest path $\pi(a,b)$ as the one having the lowest number of hops between a and b that belongs to P(a,b). Let Π be the set of the shortest paths between any pair of agents in the network. The network average shortest path is given by:

$$\bar{\pi} = \frac{\sum_i^{|\Pi|} \pi_i}{|\Pi|}.$$

Before introducing the small world coefficient itself, it is worth noting that this network is a projection from a bipartite structure; so the measurements had to be corrected by dividing them with the equivalent random graph counterparts [10]. The small world coefficient is given by:

$$\mathcal{Q} = \frac{\mathcal{CC}}{\bar{\pi}}.$$

The small world coefficient allow us to verify the connectivity and cohesion among the producers, writers and directors. The more a network exhibits characteristic of a small world, the more connected the agents are to each other and connected to agents who know each other through past collaborations. We can access the correlation between this network metric and success by associating movie's success parameters with the small world coefficient from the whole network at the time of the movie's release.

Also considering the network topology at the time the movie was released, we calculate metrics that are related to the team that produced the movie and its relative position in the network. These metrics allow to evaluate the previous experience, degree of interaction and cohesion among the agents. Let τ_m be the team that produced a given movie m with size equals to the cardinality $|\tau_m|$. Based on Uzzi [14], we define the following metrics:

(1) Average Previous Team Experience: Let τ_m^2 the binary Cartesian product of the team that produced a movie m. Let $T_E(a,b,c)$ be the number of movies produced by the agents a and b, before the current movie c. The average number of movies each pair of team members jointly produced before, considering all possible pairs in the team, i.e, the Average Team Experience, is given by:

$$\bar{T_E}(\tau_c) = \frac{\sum_{\forall(a,b)\in\tau_c^2} T_E(a,b,c)}{|\tau_c^2|}.$$

(2) Average Previous Team Shared Collaborators: Let $T_S(a,b,c)$ be the number of collaborators a pair of team

[7]Historical Currency Converter: http://currencies.apps.grandtrunk.net

[8]CPI Inflation Calculator: http://www.bls.gov/data/inflation_calculator.htm

members have in common, before the current movie c. The Average Number of Shared Collaborators is given by:

$$\bar{T_S}(\tau_c) = \frac{\sum_{\forall (a,b) \in \tau_c^2} T_S(a,b,c)}{|\tau_c^2|}.$$

(3) Average Previous Team Clustering Coefficient: Let $T_{CC}(a,c)$ be the local clustering coefficient[9] of the agent a, before the current movie c. The average previous team clustering is given by:

$$\bar{T_{CC}}(\tau_c) = \frac{\sum_{\forall a \in \tau_c} T_{CC}(a,c)}{|\tau_c|}.$$

(4) Average Previous Team Closeness: The closeness metric indicates how close a given agent is to any other agent in the whole network and it is calculated from the shortest path metric[10]. Let $T_{Cl}(a,c)$ be the closeness metric of the agent a, before the current movie c[11]. The average previous team closeness is given by:

$$\bar{T_{Cl}}(\tau_c) = \frac{\sum_{\forall a \in \tau_c} T_{Cl}(a,c)}{|\tau_c|}.$$

(5) Average Previous Team Betweenness: The betweeness metric indicates the frequency of the shortest paths from any pair of source and destination that pass through the agent a. Let $T_B(a,c)$ be the betweenness metric of the agent a, before the current movie c. The average previous team betwenness is given by:

$$\bar{T_B}(\tau_c) = \frac{\sum_{\forall a \in \tau_c} T_B(a,c)}{|\tau_c|}.$$

Focusing on the individual performance of agents, we also analyze interesting non-topological metrics. These metrics help to evaluate the individual experience and track record from members of the team.

(1) Average Previous Individual Experience: Let $I_E(a,c)$ be the number of movies produced by the agent a, before the current movie c. The average number of of movies previously produced by team members before the current movie, i.e, the Average Individual Experience, is given by:

$$\bar{I_E}(\tau_c) = \frac{\sum_{\forall a \in \tau_c} I_E(a,c)}{|\tau_c|}.$$

(2) Average Previous Team Rating: Let $T_R(a,c)$ be the average rating of the movies produced by the agent a, before the current movie c. The Average Team Rating is given by:

$$\bar{T_R}(\tau_c) = \frac{\sum_{\forall a \in \tau_c} T_R(a,c)}{|\tau_c|}.$$

(3) Average Previous Team Gross Income: Let $T_G(a,c)$ be the average gross income of the movies produced by the agent a, before the current movie c. The Average Team gross income is given by:

$$\bar{T_G}(\tau_c) = \frac{\sum_{\forall a \in \tau_c} T_G(a,c)}{|\tau_c|}.$$

[9]The set of triangles and triples are restrict to the agent neighborhood.

[10]Closeness(a) = $\frac{1}{\sum_{\forall i} \pi(a,i)}$

[11]This metric is calculated considering the whole network, but the team metric is restricted to the agents in the current movie c.

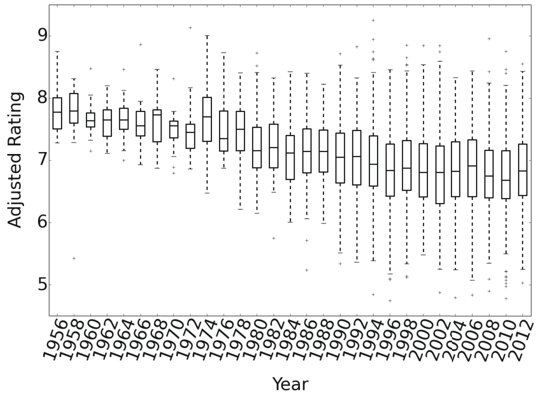

Figure 1. Movie rating distribution per year.

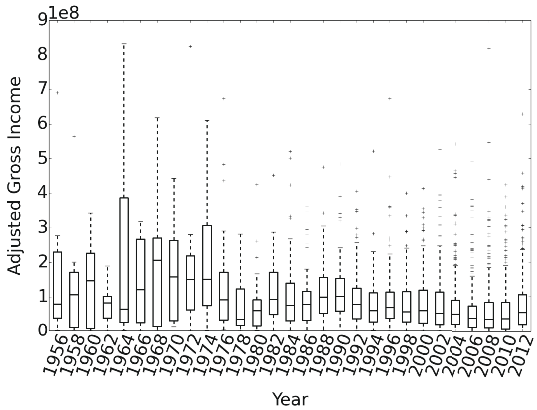

Figure 2. Movie gross income distribution per year.

V. Experimental Evaluation

Before focusing on our main analysis, we present how rating, gross income and small world coefficient evolve over time in our database. After, we discuss how topological and non-topological metrics impact the success parameters considered in our work (*rating* and *gross income*).

A. Historical Evolution of the Network

Figure 1 shows the rating distribution from 1955 to 2013. Average rating for movies decreased almost one point, in average, over the years (from ≈ 8 to ≈ 7). Interestingly, rating has been spread over the years (for instance, in 2013, the minimum rate is 5.03 and the maximum rate is 8.38). These results suggest that the average movie quality decreased over the years, from the public point of view. This effect could also be due to selection bias: possibly the bad

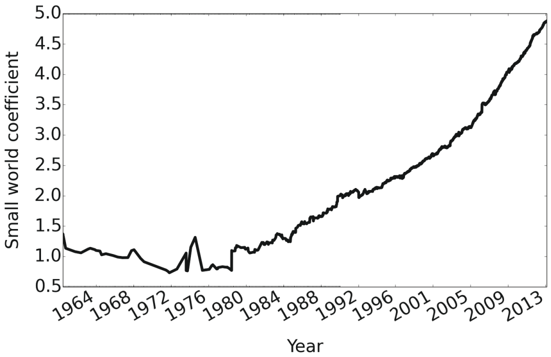

Figure 3. Evolution of the small world coefficient.

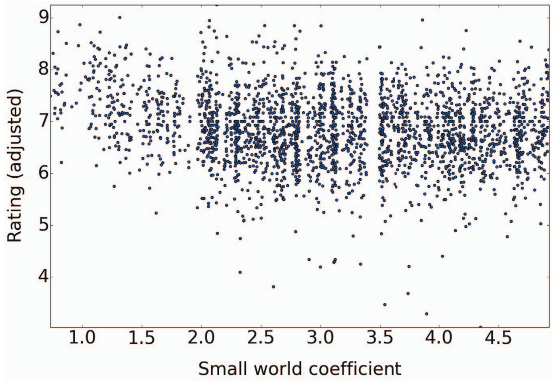

Figure 4. Rating and small world coefficient.

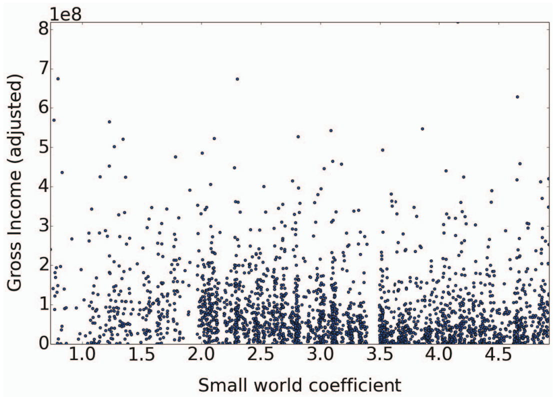

Figure 5. Gross income and small world coefficient.

movies produced a long time ago are being ignored by the public, receiving no or few ratings, whereas newer movies all receive ratings, regardless if they are good or bad.

Figure 2 depicts the gross income achieved by the movies we analyzed. Similarly to the rating, gross income also shows a decreasing pattern. Even considering only movies eligible for the TOP 250 list, there are a set of movies with much higher gross income. These results indicate the high heterogeneity of the financial success in the film industry. The movie with highest gross income in 2013 earned \$409 million.

The small world coefficient behavior is presented in Figure 3. From 1961 to 1980, small world coefficient is low. In these years, teams are very spread over the network, with very few links that do exist between them. However, since 1980 the coefficient grows monotonically, indicating high connectivity and cohesion among teams in the network. The network is getting more and more closely knit, with a large number of third-party-in-common relationships. As discussed by Uzzi and Spiro [14], the increase in the level of connections among teams can add the necessary level of credibility needed to facilitate the spread of potentially fresh but unfamiliar creative material by the producers in the network.

B. Topological Metrics

We turn our attention to better understanding how network characteristics impact the success of movies. First, we discuss the small world coefficient. Figures 4 and 5 show the results, for the rating and gross income metrics, respectively. As a global metric, its value does not depend on a specific team but on the whole network. We calculate the coefficient for a movie considering the whole network at the time the movie is released. For the movie rating (Figure 4), it is interesting to highlight that as the small world coefficient increases, the overall rating tend to decrease. There are some movies with rating below 6 (for $\mathcal{Q} > 2$). Our results tend to follow the conclusions made by Uzzi and Spiro [14] that claim that high connectivity may homogenize the pool of creative material, interfering in the production of good movies. However, it is worth nothing that we did not find the strong correlation found by Uzzi and Spiro in their work. For the gross income metric (Figure 5), small world coefficient does not reveal any tendency of correlation.

Figures 6 and 7 present the correlation between the average previous team experience and rating and gross income, respectively. First, most of the movies have low values for average previous team experience. We observe that movies with high values for this metric are less likely to receive a high rating or achieving a large gross income. We can suppose that people who always work together are less likely to have new ideas or courage to innovate. This finding agrees with many works in the literature: new collaborators are highly likely to bring new ideas, resulting in a movie with high potential of achieving success.

The results of the average previous shared collaborators ($\bar{T_S}(\tau_c)$) behavior corroborate the affirmation about the correlation between novelty and success. Figures 8 and 9 depict the results. Teams with the highest values for $\bar{T_S}(\tau_c)$ tend to think in the same way without bringing novelty to the movies that they are producing. Then, these teams tend

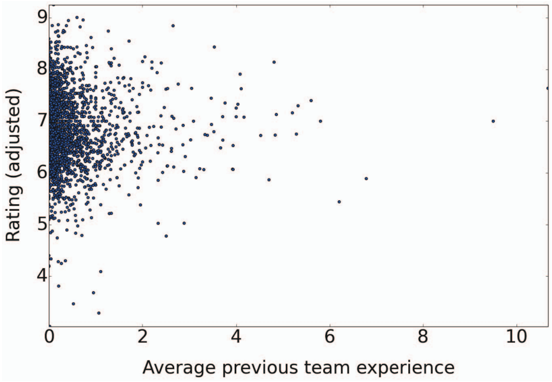

Figure 6. Rating and average previous team experience.

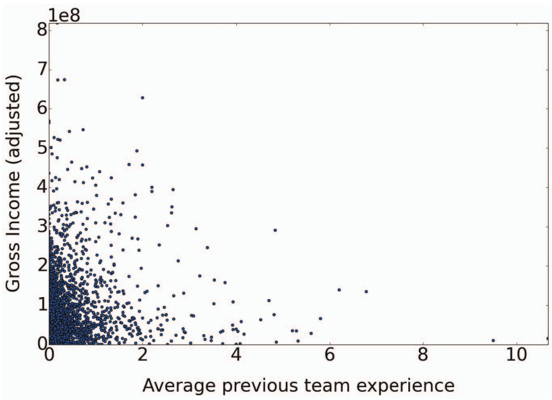

Figure 7. Gross income and average previous team experience.

to be less successful. Teams with the lowest values for the metric are the ones who generate the exceptional ratings and gross income.

Although it also represents the level of team cohesion, the average previous team clustering coefficient seems to be uncorrelated to rating or gross income, as shown in

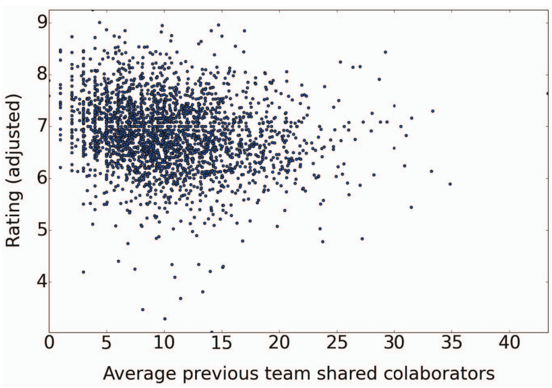

Figure 8. Ratings and average previous shared collaborators.

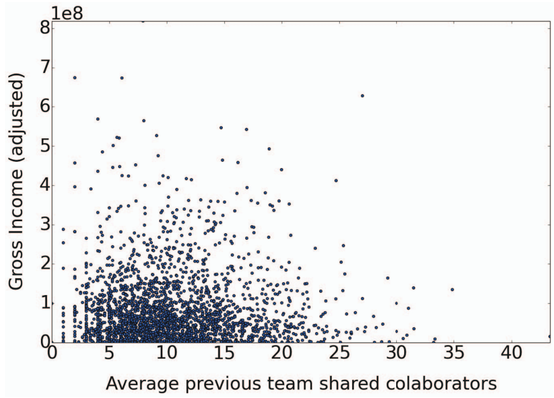

Figure 9. Gross income and average previous shared collaborators.

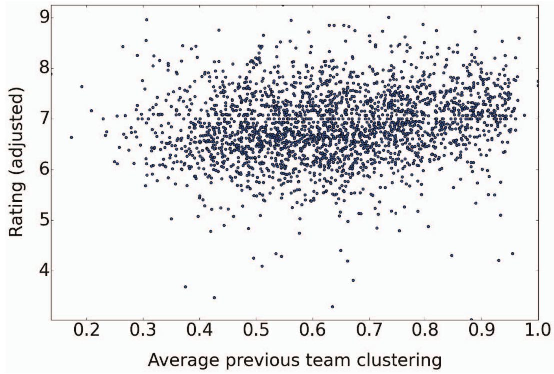

Figure 10. Rating and Average Previous Team Clustering

Figures 10 and 11. As previously discussed, it is important to have some level of previous collaboration to achieve success. However, the number of triangles does not seem to influence the movie success.

Let us focus on the average previous team closeness ($\bar{T_{Cl}}(\tau_c)$), presented in Figures 12 and 13. Movies with

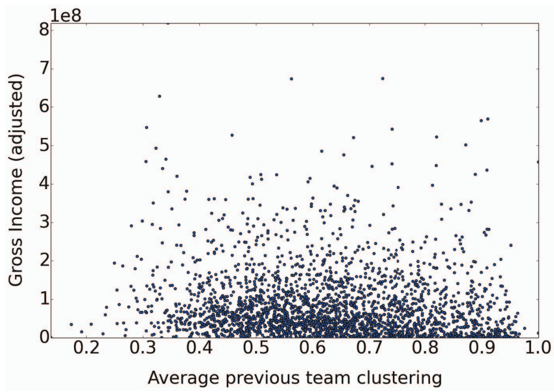

Figure 11. Gross income and Average Previous Team Clustering

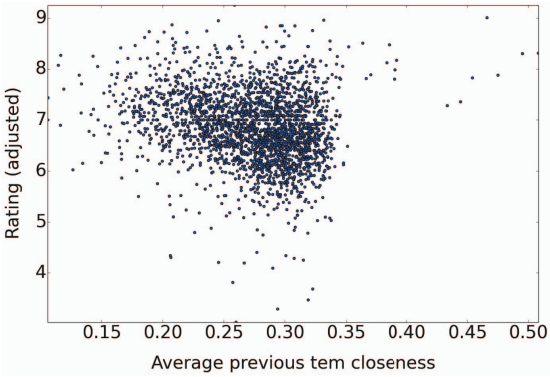

Figure 12. Rating and average previous team closeness.

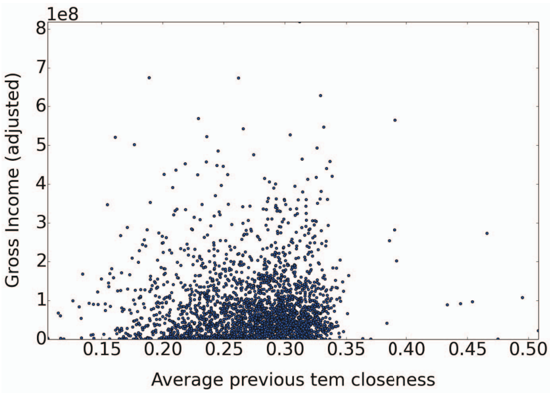

Figure 13. Gross income and averageprevious team closeness.

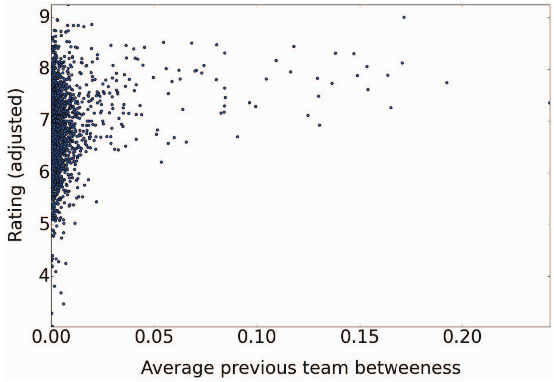

Figure 14. Rating and average previous team betweenness.

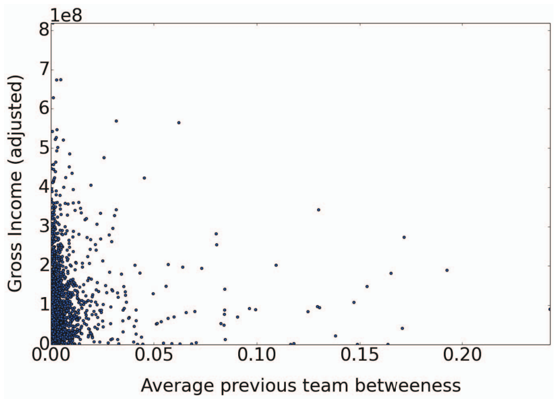

Figure 15. Gross income and average previous team betweeness.

intermediate values of $\bar{T}_{Cl}(\tau_c)$ tend to receive ratings lower than 7.0. Only some outliers with very high values for $\bar{T}_{Cl}(\tau_c)$ receive better ratings. In the other hand, teams with intermediate values of $\bar{T}_{Cl}(\tau_c)$ produce movies with amass higher gross income. We can suppose that, producers who are a few step from successful producers tend to attract the public attention inducing them to watch the movie, increasing the gross income. However, after watching these movies, the public acclamation is not that high, explaining the low rating.

Figures 14 and 15 show the results for the average previous team betweenness ($\bar{T}_B(\tau_c)$). For small values of$\bar{T}_B(\tau_c)$ the betweenness is correlated neither to rating nor gross income. However, for the rating score, values of $\bar{T}_B(\tau_c) > 0.05$ attract the rating values to values around 7.5.

C. Non-Topological Metrics

Next, we focus on analyzing the correlation between non-topological metrics and success. First, let us focus on the average previous individual experience. Movies with the highest rating scores and gross incomes tend to be produced by teams in which directors, writers and producers have little experience in the past. Figures 16 and 17 show the results. Interestingly, teams with much experience in the past tend to have less success. This result counter intuitively shows that teams that already produced many movies before in fact are less likely to produce movies with high public acclamation and high gross income. We may suppose that teams with less experience are mostly composed by young people who are not afraid to innovate. Of course, we can not generalize our conjecture. There are many producers with large experience who frequently produce movies that achieve tremendous success.

Average previous team rating is the metric that best correlates and explains movies' rating and gross income. Figures 18 and Figure 19 present the results. There is a clear correlation between the metric and the current rating. Moreover, most of the movies with amassed the highest gross income were produced by teams with average previous rating above 6.0. Average previous team gross income, instead, seems do not be correlated either to the movie's rating or gross income, as shown in Figures 20 and 21.

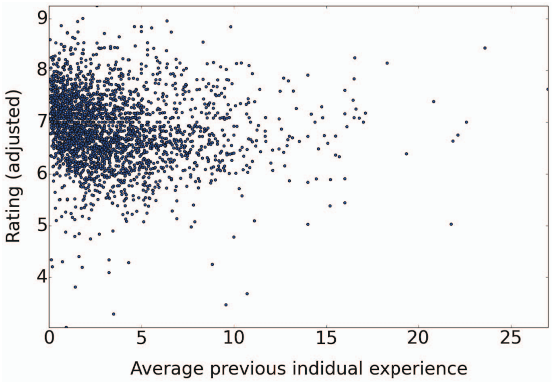

Figure 16. Rating and average individual experience

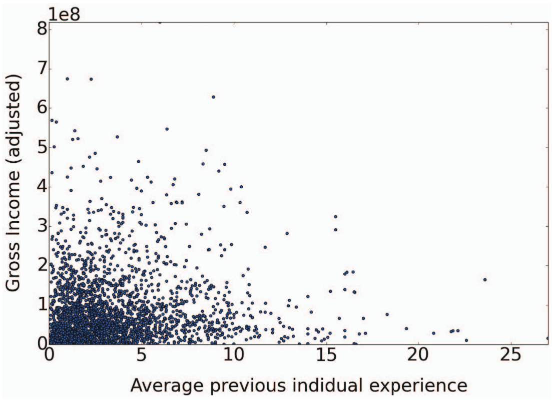

Figure 17. Gross income and average individual experience

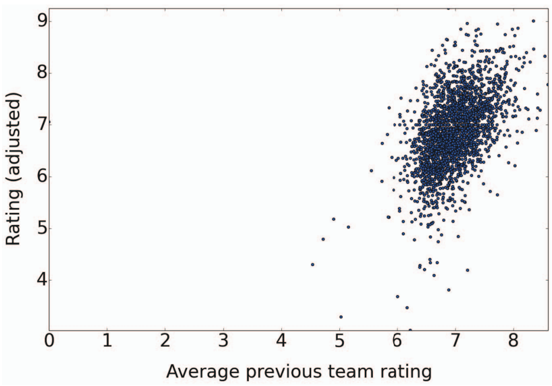

Figure 18. Rating and average previous team rating

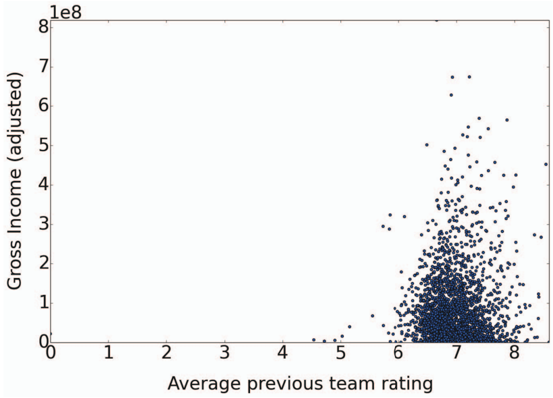

Figure 19. Gross income and average previous team rating.

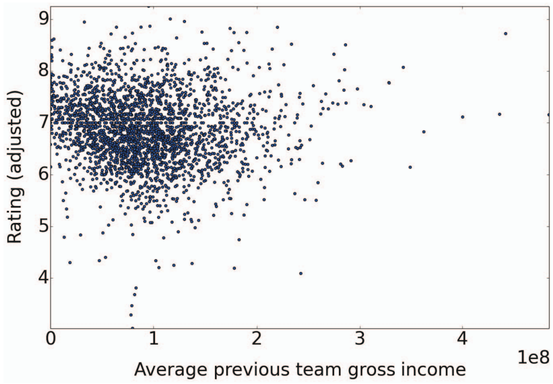

Figure 20. Rating and average previous team gross income.

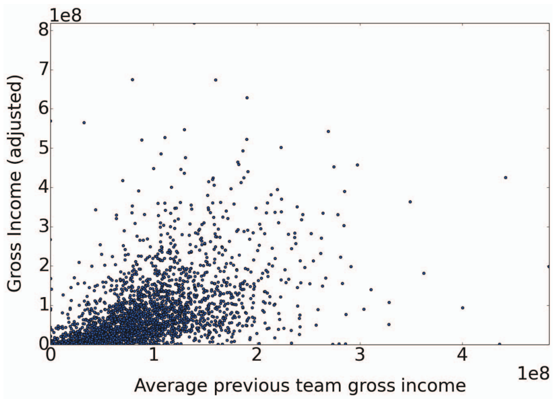

Figure 21. Gross income and average previous team gross income.

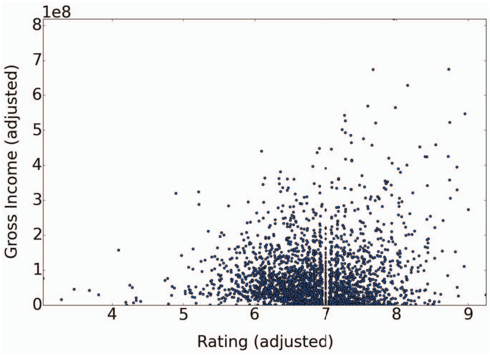

Figure 22. Rating and gross income correlation.

D. *Rating versus gross income*

An interesting question to explore is whether movie's gross income and ratings are correlated. Figure 22 shows that movies with rating above 6.0 tend to achieve the highest gross incomes in our database. Then, rating and gross income tend to be correlated considering movies in the data we analyzed.

VI. Conclusion

We have presented a broad study on how topological and non-topological metrics of the network of directors, producers and writers impact the success of a movie produced by this team of people. Our findings are very interesting. Non-topological metrics, such as team's average previous rating, centered on the individual or on the team itself had more clearly correlation to the success metrics. Some topological metrics showed to be weakly correlated to the movie success. Interestingly, we found that teams with too much past experience perform worse than teams with fresher agents, reinforcing the assumption that novelty helps to form successful teams.

Besides giving some insights of the correlation between team formation and success, our results are important to show that the team success in film industry is not that simple to characterize, and more elaborate metrics have to be considered. For instance, we believe that the team success can be explained by considering jointly individual and team characteristics or by a more elaborate combination of topological metrics. As an improvement, other strategies can be used to aggregate value from the team members, for example the maximum value or the harmonic mean. Furthermore, other network metrics can be used, such as Burt's structural hole index. We plan to address such points in future work. Furthermore, we are working on employing more metrics to measure movie success, including popularity in online social networks and ratings from other websites, such as Metacritics and Rotten Tomatoes.

Acknowledgments. This work was partially funded by CAPES, CNPq and FAPEMIG.

References

[1] R. S. Burt. Structural Holes and Good Ideas. *The American Journal of Sociology*, 110(2):349–399, 2004.

[2] Z. Chen and J. Guan. The impact of small world on innovation: An empirical study of 16 countries. *Journal of Informetrics*, 4:97–106, 2010.

[3] A. Elberse. The Power of Stars: Do Star Actors Drive the Success of Movies? *Journal of Marketing*, 71:102–120, Oct. 2007.

[4] J. Kleinberg, S. Suri, É. Tardos, and T. Wexler. Strategic network formation with structural holes. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 284–293. ACM, 2008.

[5] E. Y. Li, C. H. Liao, and H. R. Yen. Co-authorship networks and research impact: A social capital perspective. *Research Policy*, 42(9):1515–1530, 2013.

[6] K. Nemoto, P. Gloor, and R. Laubacher. Social capital increases efficiency of collaboration among wikipedia editors. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, pages 231–240. ACM, 2011.

[7] M. Newman. The Structure and Function of Complex Networks. *SIAM Review*, pages 167–256, 2003.

[8] M. E. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E*, 64(1), 2001.

[9] M. E. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.

[10] M. E. Newman, S. H. Strogatz, , and D. J. Watts. Network effects: Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E Stat. Nonlin Soft Matter Phys*, 64(2), 2001.

[11] M. A. Schilling and C. C. Phelps. Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science*, 53(7):1113–1126, 2007.

[12] P. V. Singh, Y. Tan, and V. Mookerjee. Network effects: The influence of structural capital on open source project success. *MIS Quarterly*, 35(4), 2011.

[13] D. Stokols, K. Hall, B. Taylor, and R. Moser. The science of team science: overview of the field and introduction to the supplement. *American Journal of Preventive Medicine*, 35(2):77–89, 2008.

[14] B. Uzzi and J. Spiro. Collaboration and creativity: The small world problem. *American journal of sociology*, 111(2):447–504, 2005.