Companionship Network Analysis: Information Gain Ranking, Time Dynamics and Bipartite Properties

Lukas Dennert
Ludwig Maximilian University
Munich, Germany
University of Porto
up202110822
lukas.dennert@campus.lmu.de

Matheus Teixeira
Federal University of Espírito
Santo, Alegre, Brazil
University of Porto
up202110548
matheus.b.teixeira@pm.me

Antonia Halbig
Ludwig Maximilian University
Munich, Germany
University of Porto
up202110759
a.halbig@campus.lmu.de

ABSTRACT

Bipartite networks allow for an exclusive interplay between the two subsets of nodes, such as rankings based on the connections. We propose a novel ranking method to rank nodes depending on the received reviews in conjunction with a weighting within the opposite subset of nodes prioritizing information gain over mere quantity. This ranking is applied to seller nodes of a bipartite, dynamic, signed multigraph of escorts on the basis of the corresponding buyers evaluating their services. Supplementary, we analyze the development of nodes and ratings over time and provide some insights on the respective projections.

1. INTRODUCTION

Solely ranking items by the quantity of some metric does not ensure an informative outcome. Having entities rating a purchase provides a base to rank the service or product they bought. But if these entities tend to always give the same rating, their reviews lack in informative value, as it can be assumed they will not change the rating even if it the quality of the product or service drops [Zhang, 2011]. With the unique property of bipartite networks one subset of nodes can be used to rate the other one, inspired by the hubs and authorities concept of Kleinberg [Kleinberg, 1999]. First we give a broad overview over the escort network from Rocha et al. [Network, 2010] and then describe the design of the ranking in section 3 applied to the same network. Also having the timestamps of the edges to our disposal opens up possibilities to outline various trajectories and developments over time, which are presented in the sections after.

2. GENERAL NETWORK ANALYSIS

The chosen network is represented by a bipartite multigraph, where signed edges are constructed over time. It consists of 16730 nodes and 50632 edges between them, containing one giant component which will be in the main focus in further analyses.
In this bipartite network the two types of nodes are categorized as buyers, who buy a service, and sellers, who sell this service to the buyers. Buyers do not interact with each other and sellers do also not have edges between themselves, thus the property of being bipartite and making it suitable for projections of only one category. Nevertheless, as it is a multigraph, it is possible to have multiple edges between two nodes, indicating that a buyer bought multiple times from the same seller. In the code nodes are treated similarly with the only distinction that the names of buyer-nodes start with a lowercase 'b' and the names of seller-nodes start with a lowercase 's'.
The edges originally represent an escort service being sold from a female escort to a prospective male buyer. In this case the service has already been bought and finalized, so it is a retrospective review from the buyer side about the performance of the seller side. The review is represented in an edge weighting of either '-1' for a negative, '0' for a neutral and '+1' for a positive experience. A prominent feature of these edges is their temporal factor. The reviews are given over a time window slightly bigger than 6 years, from the 31 of August 2002 until the 10 of October 2008. With the inclusion of a timestamp on the edges the network dynamics over time can be examined, which will be further addressed in one of the subsequent chapters.
The bipartite property of the graph, in addition to the timestamp and the review, allows for an analysis of the interplay between the different node types. Section 3 outlines multiple rankings of buyers and sellers, also in dependence of each other.

2.1 Network Components

The network consists of 418 unlinked components, whereof smaller components are examined coarsely and the main analysis is made on the giant component. All components of the network are partitioned as follows: 357 of size 2, 44 of size 3, 11 of size 4, 2 of size 5, 2 of size 6, 1 of size 8 and 1 of size 15810. We take a quick look into smaller network components, before coming to an analysis of the giant component.

2.1.1 Small Components

All 401 components of size 2 and size 3 carry no substantial information, as they are either too small, have too few edges or all their edges are one-sided. Same goes for most components of size 4, with one exception, which consists of four edges between two buyers (here called 'b1' and 'b2') and two sellers (here called 's1' and 's2'). The chronological order would be: 1. b1 bought from s1 and left a review of '-1', 2. b1 bought from s1 again and left a review of '-1' again, 3. b1 changed seller, bought from s2 and left a review of '0' and lastly, 4. b2 bought from s2 and left a review of '-1'.
If we would spin a little story around this component it

could look like the following: First b1 bought the service of s1 and was not happy with what he got. He chose to gave s1 another chance and bought again from s1. But because he was not satisfied again, he switched seller and bought from s2 the next time. As the service he received was better from s2 he passed this information on to b2, who then also turned to s2 and bought a service there. Apparently b2 did not like the recommendation, as he rated the service as insufficient ('-1').
To the components of size 3 it can also be said, that 26 have two buyers and one seller, whereas the remaining 18 have two sellers and one buyer. For the components of size 4 there are 2 with three buyers and one seller, 6 with three sellers and one buyer and 3 balanced ones with two buyers and two sellers. All in all not carrying too much information, as they seem quite evened out and, because of their size, do not allow for a more profound perception.
Components of size 5, 6 and 8 have not been examined any further yet. We consider the giant component to be the most important part of this network, as it includes 94.5% of all nodes and 98.98% of all edges.

2.1.2 *Giant Component*

The giant component contains most information and with the temporal attribute of the weighted edges a lot of different information can be discovered.
It contains a total of 15810 nodes, where 9652 are buyers and 6158 are sellers, making the entire component more buyer-heavy with an approximate ratio of 1.5 buyers per seller. The nodes are connected by 50116 edges in total and therefore an average node degree of 6.34. 41672 of the reviews are positive ('+1'), whereas only 4148 are neutral ('0') and 4296 are negative ('-1'), retaining a ratio of roughly ten times more positive reviews than neutral or negative ones or a total of 83.15% of positive reviews.
Generally speaking positive reviews are more common in the companionship network, but so are they in other online review platforms. This trend of the review ratio is close to the numbers of the worlds largest online retailer amazon, where products can be reviewed and rated. They had 87% positive reviews in the U.S. in April of 2020, with slightly fluctuating, but still stable in this region, numbers in the moths before [Kaziukenas, 2020].
With the buyer-nodes, the seller-nodes and the connecting reviews a multitude of rankings can be constructed, which will be further explained in the following section.

3. RANKING NODES

All nodes can be ranked according to the reviews they are associated with, their node degree and even the ranking of nodes that they are linked to, similar to hubs and authorities. In order to reduce clutter for these rankings and cut off the long tail of the distribution, we only rank nodes with degree three or more. For both buyers and sellers we created a straightforward ranking according to the rankings they either gave or received. Additionally more sophisticated rankings have been constructed, with the focus on finding a more meaningful ranking for the sellers. Therefore we had to first construct a buyer ranking including their overall review sentiment, as buyers who mainly give either positive or negative reviews only add little informational value to the reviews of sellers. Dependent on the sophisticated buyer ranking we then designed an expressive ranking for the sellers, where the reviews of buyers are weighted according to their new rank.

3.1 Ranking Buyers

All 4200 buyers are ranked in two different ways, at first a plain ranking and subsequently a ranking depending on the given reviews. The plain ranking takes the average over all reviews the buyer has published and ranks them accordingly. If two buyers have the same average review they are compared against the amount of reviews they gave as a secondary dimension. The more reviews, the better the ranking. This can be seen at the top of the ranking, where every node has an average review of '1.0' revealing that they only gave out positive reviews. For all buyers with the same average review, the one with 1776 reviews is ranked in first place and the one with 1521 reviews is ranked in second place.
This already leads to the issue we found with this ranking. Even though the first node handed out a lot of reviews and actively produces data for the network, all of his 1776 reviews are positive and therefore do not provide a lot of added information. In order to counter this problem, we introduce the second ranking, which also takes the amount of same reviews into account. The new ranking calculates every buyer rank (R_i^B) according to the following formula:

$$R_i^B = (1.1 - |avg\_review|) \cdot \log_2(num\_reviews) \quad (1)$$

The calculation consists of two parts. First the average of all given reviews are taken into account and secondly the amount of given reviews. The absolute of the average of all given reviews is taken and subtracted from the fixed number 1.1. This is done to reduce the importance of buyers who have an average review of either '1', '-1' or close to one of those two, as their reviews mainly return the same positive or negative result, thus adding no information to the actual quality of the sellers. As we want to reward active usage of the recommendation system, we included the amount of total given reviews into the calculation. Nevertheless it is dampened by a logarithmic function, as the sheer amount of reviews itself should not give a buyer linearly more voting powers.
With this updated ranking method we ranked all buyer-nodes again and used the newly acquired ranking for a more sophisticated seller ranking.

3.2 Ranking Sellers

In order to observe the change in rank for all 2848 seller-nodes, we also had to construct a plain and a more sophisticated ranking. The first ranking is similar to the plain ranking of buyers. All seller-nodes are first ranked according to the average review they received and as a second metric by the amount of reviews they received. Here the same problem arises as in the buyer ranking, where highly or lowly ranked nodes do not add a lot of information, as they always give the same reviews. So we also constructed a more complex seller ranking in dependence of the more sophisticated buyer ranking, calculating the new ranks of sellers (R_i^S) as per this equation:

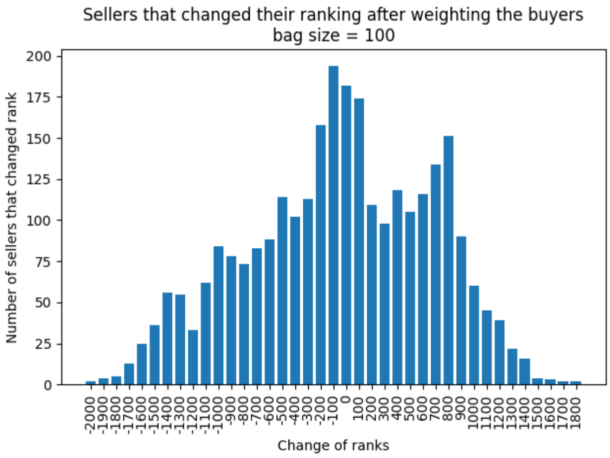

Figure 1: Change of sellers from the plain to the more sophisticated ranking, allocated in bags of size 100 for improved readability. Negative numbers signify a drop in ranks and positive numbers a climb.

$$R_i^S = \sum_j R_j^S \cdot review_j \qquad (2)$$

The new rank of a seller (R_i^S) is determined by the sum of all reviews received multiplied by the updated weighting of the buyer who gave this specific review. The index j iterates through all edges, and therefore reviews, connected to buyers and to the specific seller i.
Comparing the updated ranking with the plain seller ranking, we can notice drastic changes in position. These changes are visualized in Figure 1, where a negative number represents a drop in ranks from the plain ranking and a positive number a rise.
Most sellers fall within the bags between '-200' and '100' implying that most nodes do not undergo a strong change in ranks, as they were all quite close already in their initial ranking. A few nodes on the other hand seem to undergo a dramatic change, as they fall in the range from '-2000' to '-1400' or from '1400' to '1900', which is a change in position of half the ranking either down or up. Here it is interesting to note that there are nodes dropping by 2000 positions, but there is no node climbing up the same amount. The maximum rise has been denoted as an increase of 1900 ranks. Furthermore a slight rise between '600' and '800' can be observed, equalling out the otherwise uniformly higher columns on the negative side.
Figure 1 indicates the importance of a more elaborated ranking system than just the base average, as there are many nodes changing their position and even some nodes with radical jumps in the ranking. Not only do we provide a novel ranking system for this network, we also looked further into timely aspects of the nodes and the change the network undergoes over time.

4. OVER-TIME ANALYSIS

Given the timestamps present in the edge information, looking at the network's characteristics and their development throughout its growth over time proved to be a good approach for an in-depth analysis.
From the insights we gathered with this analysis on the network, we came up with speculation on the website's social interactions and dynamics [Rocha, 2010].

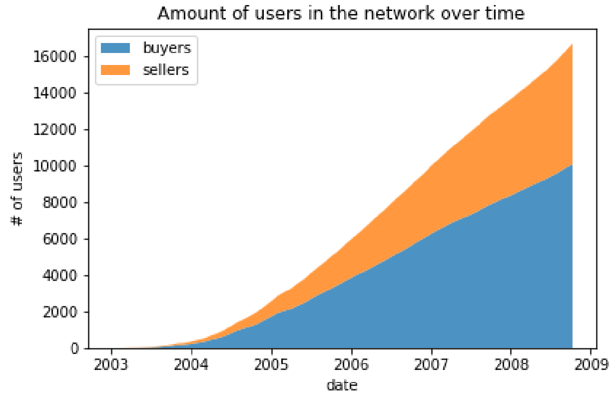

Figure 2: Amount of individual users reviewing at least once or being reviewed at least once.

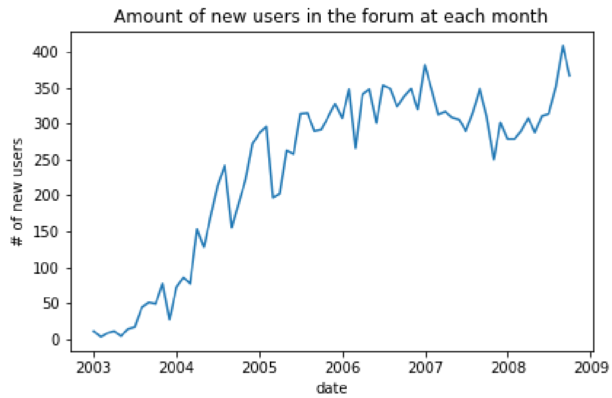

Figure 3: Amount of new users that took part in a review, as buyers or sellers, for the first time at each month.

4.1 Network Growth

At figure 2, a linear growth can be clearly noticed for both node types after a short polynomial ascendance, which could be the consequence of a initial user "boom" following the website birth and a later stabilised growth of its user base. Figure 3 shows the growth in a monthly perspective, corroborating the previous statement.
On the other hand, when looking at figure 4, we can see that the amount of monthly new reviews keeps growing. This goes along with what is expected for when the user base is in constant increase, and shows that users do not fall inactive after their first reviews. If that was the case, we would expect a graph with similar behavior to figure 3.
Something interesting happens in figure 5, when we plot the monthly amount of active users and new reviews together. There is a clear correlation between the two measurements, which could be explained by a sporadic characteristic on the

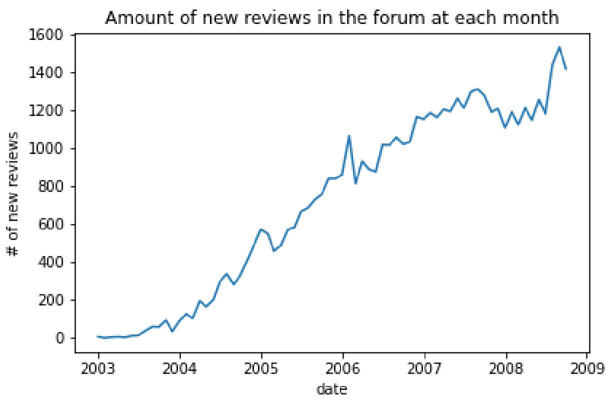

Figure 4: Amount of new reviews in the forum at each month.

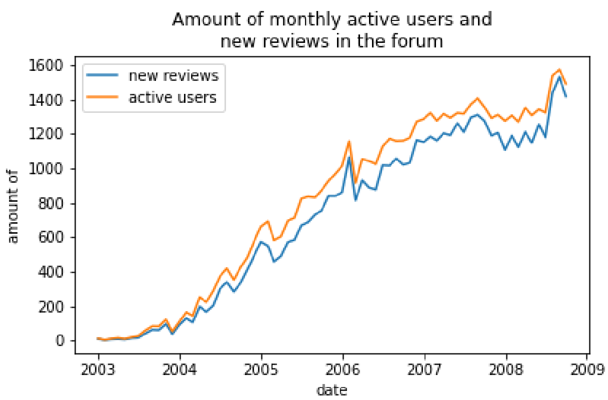

Figure 5: Comparison of new reviews and active users per month.

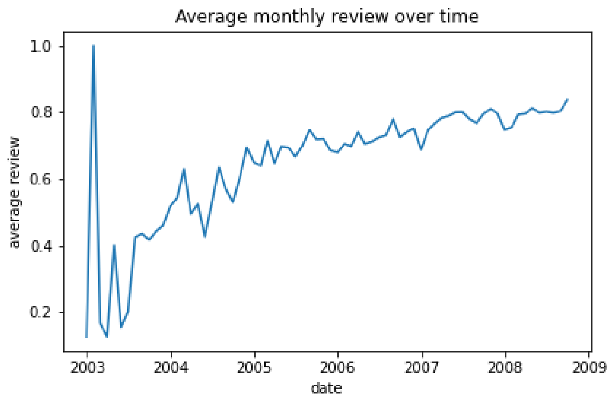

Figure 6: Mean review value of buyer's reviews on sellers over time.

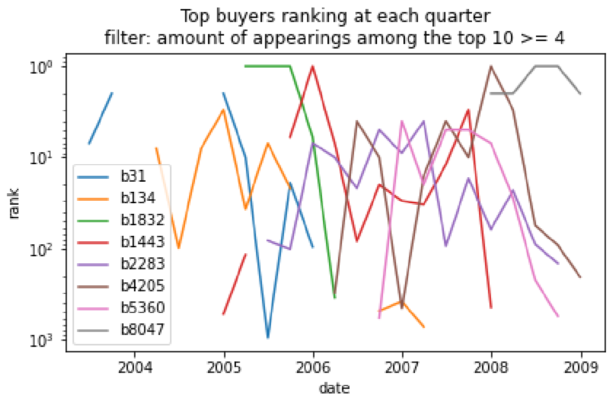

Figure 7: Top buyers ranking over time, with α=4 and β=10 as parameters for the user codes selection.

hiring from the individual buyers — in other words, most buyers not hiring sellers very often.

4.2 Review Trajectory

At figure 6, an initial instability can be explained by the small amount of users in the website, followed by a convergence towards a high positive score.
This could mean that the website worked as intended, providing to the users a reliable source of recommendations and resulting in a high satisfaction rate.

4.3 Top Buyers and Sellers

For figure 7 and figure 8, we took the user codes of some of the forum's best buyers and sellers and plotted their ranking progression over time.
Ranking here is not to be confused with section 3's definition for the same name. In this case, it means the user's position on a rank of how many reviews they are linked to of each quarter.
The users that appear on these plots are chosen by defining an arbitrary number of times α they should appear among the top β buyers or sellers.
At figure 9, we can see that if we apply the same α and

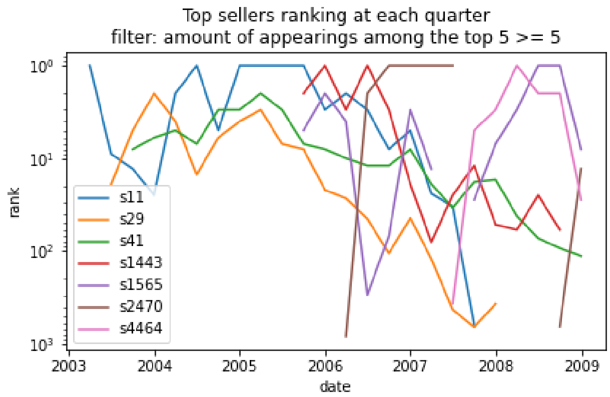

Figure 8: Top sellers ranking over time, with α=5 and β=5 as parameters for the user codes selection.

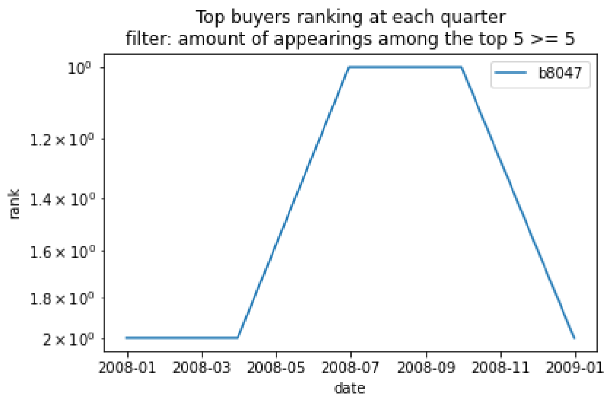

Figure 9: Top buyers ranking over time, with α=5 and β=5 as parameters for the user codes selection.

β parameters we used for the sellers to buyers we will have only one user left. This indicates that there are more sellers that stay among the ranking's top for longer periods of time.

5. BIPARTITE PROJECTIONS

In this section we will take a look at the projections of the bipartite network, once projecting the buyer nodes and the other time the seller nodes. The projections are created using Gephi[1] and the Plugin Multimode Networks Transformations. This tool was used on the giant component of the complete network, neglecting separated smaller components. The giant component includes the wast majority of the complete network, consisting of approximately 95% of its nodes.

5.1 Projection Overview

The buyer projection consists of 9652 nodes and 668142 edges, without any parallel edges. Since the number of edges is very high the projections results in a high average degree of 138.446.
The seller projection is made up of 6158 nodes and 183383 edges. It also does not include parallel edges. The resulting average degree is lower than in the buyer network with a value of 59.559.
A more detailed overview over the statistical data of the networks can be found in the table 1.

	Giant Component	Buyer Projection	Seller Projection
Nodes	15810	9652	6158
Edges	50116	668142	183383
Avg. Degree	3.17	138.446	59.559
Network Diameter	17	8	8
Avg. Path Length	5.78	2.77	3.05
Modularity	0.676	0.412	0.621
Communities	16	15	16

Table 1: Table containing statistical data of the giant component of the network, the bipartite projection on the buyer nodes and the bipartite projection on the seller nodes.

5.1.1 Average Degree

In the complete bipartite network the degree of seller nodes strongly outweighs the ones of buyer nodes. This can be explained by the fact, that buyers make 60% of the nodes of the network. Due to the bipartite nature of the network, they can only connect with seller nodes, which there are less of. The sellers therefore end up having a higher degree on average. When looking at the projections, the contrary is visible. The average degree of the buyer projection is 138.446, while in the seller projection nodes only have a degree of 59.559 on average. This can be explained by the degrees of the node types of the complete network. Since

[1] https://gephi.org/

the sellers have a higher average degree, they create more edges between the nodes in the buyer projection.

5.1.2 Comparing Importance Metrics

Nodes can be ranked based on different measurements of importance. Besides the degree each node can be evaluated by betweenness centrality C_B, closeness centrality C_C or PageRank amongst others. When comparing these measures for nodes of the giant component of the complete network as well as each projection, some similarities and differences are visible.
All networks have most of their highest degree nodes located in the big central cluster. Each of the other clusters has some higher degree nodes as well, which seem to have a central role within that community.
Looking at betweenness centrality, the important nodes are more diversely located within the networks. The projections only have a few nodes with high C_B. In the giant component of the complete network the majority of nodes with high C_B values are sellers with central positioning in the biggest cluster in the layout.
PageRank behaves very similar to the degree measurement. When ranking nodes on closeness centrality most of the nodes in the projections end up with similar values around 0.3 and 0.4. The giant component of the complete network has most nodes in values around 0.2 to 0.3.

5.2 Communities

This section analyses the resulting communities after running a modularity algorithm in Gephi on the giant component of the complete network as well as the projections created in the section before.

5.2.1 Comparing Communities with Projections

The giant component of the complete network is divided into 16 partitions. Using the Force Atlas 2 layout algorithm in Gephi, each of these communities is clearly separated from each other. There is a central cluster in the middle of the network, which takes up over 30% of the nodes of the graph. In it lie the nodes with highest degree and centrality values. This cluster is separated into multiple communities, with the two biggest communities being part of it.
Looking at the visualisation of the communities of the network in Figure 10 this is visible. For better comparison, each visualisation uses the same layout, meaning seller nodes in the projection have the same position as in the giant component, and the same for the buyer nodes.
The buyer projections results in 15 communities. Compared to the giant component network some of the separated communities are summed up here. An example of this is observable when comparing Figure 10 and 11. The bright green community in the latter network approximately combines the bottom left communities (shown in green and blue) of the original network. Since the number of communities is still 15, this is balanced by the fact that three other communities only consist of very few nodes.
The seller projections results in 16 communities. Similar to the buyer network before, some communities are a sum of multiple approximate communities of the giant component network. Looking at Figure 12, this mostly affects the big central cluster. The bright blue community consists of 30%

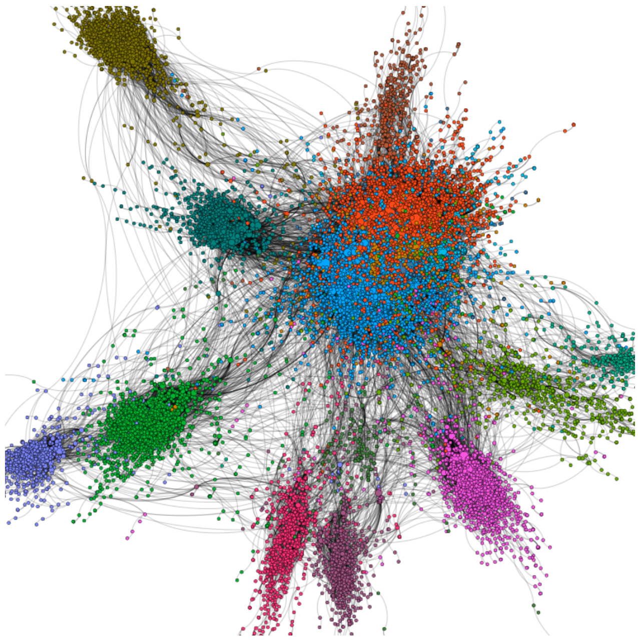

Figure 10: The figure shows a central cut-out of the visualization of the giant component of the complete network. Each color relates to one of the communities. The size of each node represents its degree, with a higher degree resulting in a larger size. The layout is created by using Force Atlas 2 on the network in Gephi.

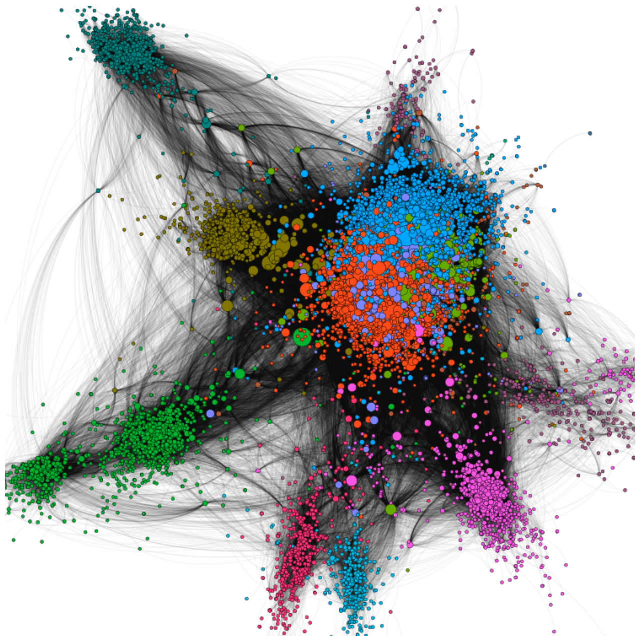

Figure 11: The figure shows a central cut-out of the visualization of the bipartite projection on buyer nodes. Each color relates to one of the communities. The size of each node represents its degree, with a higher degree resulting in a larger size. The layout is the same as of Fig. 10, i.e., each buyer node has the same location in both images.

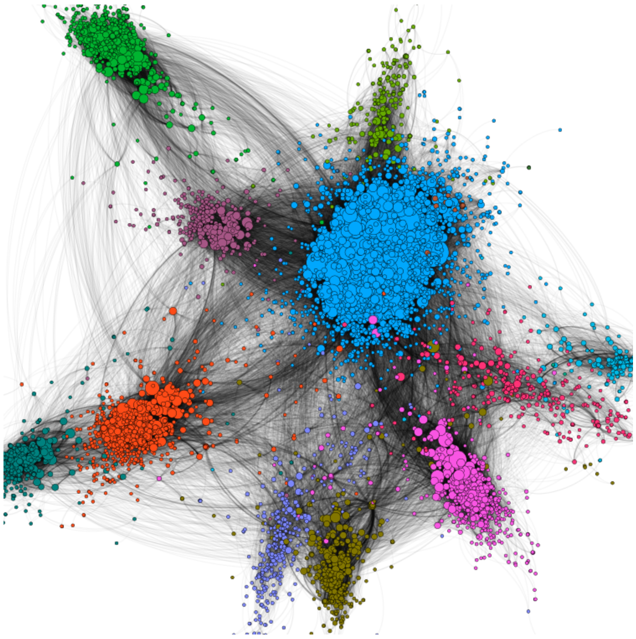

Figure 12: The figure shows a central cut-out of the visualization of the bipartite projection on seller nodes. Each color relates to one of the communities. The size of each node represents its degree, with a higher degree resulting in a larger size. The layout is the same as of Fig. 10, i.e., each seller node has the same location in both images.

of the nodes of the projection and is a summarized version of multiple central partitions of the complete network. To balance the combined communities, there are five communities with only a few nodes.

5.2.2 *Modularities*

Modularity measures how well a network is partitioned into communities. It ranges from -1 to 1, with a value above $0.3 - 0.7$ implying a significant community structure. Looking at our results, the giant component of the complete network has the highest modularity of 0.676. The seller projections results in similar modularity of 0.621, while the buyer projection value is only 0.412. Still all of them prove a significant community structure, with the buyer projection having the worst resulting partitions.

5.2.3 *Semantic Emergence of Communities*

The original data set only provides information about the review, its creation date and target and source of this interaction. Therefore, we have no insight into the semantic background of the communities. The network can be compared to a social network, since the data was collected from reviews and interactions between buyers and sellers in an online forum from Brazil [Rocha, 2010].
With this comparison some possible factors for the division come to mind. Communities could originate in groups of different interest and type of interaction. Buyers with similar sexual interest or fetishes end up reviewing the same sellers and interacting with each other on the forum, where these reviews are communicated. Another possible source of separation is location, i.e., buyers are more likely to review sellers from the same area and are less willing to travel far for a certain service.

6. CONCLUSIONS

Depending on the objective of a ranking and the condition of a bipartite network, we recommend using the introduced ranking by us to maximize information gain. Our comparison to the plain quantitative ranking highlights the drastic change in ranks and therefore also in information gain.
We also conclude more generally that setting up the recommendation system for the network was a success, as can be seen by its constant growth and increasing amount of active users. Seeing these parameters increase meets the expected behavior for a social website and therefore also an underlying social network. This growth can be seen by the always positive and increasing average rating on the platform, which seems to reflect an overall satisfaction among the user base.
Furthermore, by looking at the created partitions of the network and its projections, a clear community structure is visible, which is reflected in the modularity values. These communities show correlations to those created by social networks, with reasons for the partitioning being different interest groups or rooted in geographical location.

7. REFERENCES

Luis E. C. Rocha and Fredrik Liljeros and Petter Holme, ia-Escorts-Dynamic, `https://networkrepository.com/ia-escorts-dynamic.php`

Juozas Kaziokenas, 2020, Marketplacepulse, "One Million Unhappy Amazon Shoppers", `https://www.marketplacepulse.com/articles/one-million-unhappy-amazon-shoppers`

Richong Zhang and Thomas Tran, 2011, Knowledge and Information Systems, "An information gain-based approach for recommending useful product reviews"

Jon M. Kleinberg, 1999, Journal of the ACM (JACM), "Authoritative sources in a hyperlinked environment"

Luis E. C. Rocha and Fredrik Liljeros and Petter Holme, 2010, Proceedings of the National Academy of Sciences, "Information dynamics shape the sexual networks of Internet-mediated prostitution", `https://www.pnas.org/doi/abs/10.1073/pnas.0914080107`