Social Network Analysis of DISI Publication

Lucrezia Ilvento, Computer Science, 0001127610
Erik Koci, Computer Science, 0001136721
Valerio Desiati, Computer Science, 0001138895

1 Introduction

Social Network Analysis (SNA) is a methodological framework used across various fields, such as the study of the structure, relationships, and interactions within complex systems. This project applies SNA techniques to the Academic collaboration network of Professors at the Department of Computer Science, at the University of Bologna. Academic networks serve as a crucial platform for knowledge exchange so we'll focus on authorship patterns extracted from publication records.

The main objectives of this project include:

- Identifying central figures in the network based on various centrality measures.
- Detecting collaborative communities and analyzing their composition.
- Evaluating network robustness through resilience and density metrics.
- Assessing structural characteristics like clustering coefficients and assortativity.

Data extracted from academic publications and co-authorship records are processed into graphs for an analysis. The findings provide insights into influential individuals, tightly connected groups, and the overall network structure, which are visualized through comprehensive metrics and plots.

2 Problem and Motivation

Collaboration is fundamental to academic research, enabling the exchange of ideas. However, understanding the structure and dynamics of these collaborations can be challenging, particularly in large academic departments.

The main motivation behind this analysis is to:

- Identify key contributors who play central roles.
- Understand how different subgroups or communities of researchers interact, helping to promote interdisciplinary work and identifying potential areas for improvement in collaboration.

- Analyze network resilience under different scenarios, such as targeted removal of key nodes or random failures, to provide insight into the robustness of the network.

The impact of this analysis lies in both theoretical and practical contributions. Theoretically, it enhances our understanding of collaboration structures in academic networks using SNA techniques. Practically, it provides actionable insights for the department to improve collaboration, optimize resource allocation, and strengthen the robustness of its research ecosystem.

3 Datasets

The data for this project was primarily sourced from publicly available academic publication databases [2], focusing on the research output of faculty members from the Department of Computer Science and Engineering at the University of Bologna. The core dataset consists of metadata related to scientific papers published by the department's faculty, including paper titles, authors, and the journals or conferences in which the papers were published.

To render the data in a format suitable for analysis, we have developed a web scraper [4]. Python was the primary programming language and libraries such as `BeautifulSoup` were used to extract relevant publication information from academic databases. The raw data, which included paper titles and contributor details, were collected and stored in CSV files.

For data handling and manipulation, `Pandas` [3] was employed for data cleaning and structuring, ensuring the dataset was correctly formatted for analysis.

`NetworkX` [1] was used to construct and analyze the co-authorship network, providing tools for computing centrality measures and community detection metrics.

The dataset includes a comprehensive list of publications for each faculty member, which was used to construct the collaboration network. The co-authorship network analyzed in this study is a monomodal, undirected graph, where:

- Nodes represent faculty members in the Department of Computer Science, University of Bologna.
- Edges indicate co-authorship relationships between two Professors, based on their joint publication of a paper. Edges are undirected, reflecting that co-authorship is mutual and bidirectional.

4 Validity and Reliability

Ensuring the validity and reliability of the model used in this study is essential for drawing meaningful insights into the academic collaboration dynamics.

Validity

The dataset used in this study is based on data gathered from publicly available academic sources by the university's institutional page. Centrality measures and community detection algorithms applied to the co-authorship network offer valuable insights into the

interactions among faculty members, serving as a reasonable approximation of research collaborations.
The dataset primarily focuses on published works, which are the most reliable documentation of scientific collaboration. As such, the model provides a credible representation of the academic network.

Reliability

Reliability, or the consistency and reproducibility of the study's findings, was ensured through a rigorous data processing workflow.
The study's reliability is further supported by the accessibility of the source data, ensuring that others can independently verify the findings or apply the methodology in different contexts or departments.

5 Measures and Results

5.1 Measures

To analyze the collaboration network, we applied several Social Network Analysis key measures. Each measure was selected for its ability to highlight specific structural and dynamic aspects of the network. The measures used are:

- **Degree centrality:** This metric quantifies the number of direct connections (co-authors) a node (Faculty member) has. Faculty members with high degree centrality frequently collaborate with others, serving as hubs within the network. Identifying these individuals highlights researchers who play a critical role in fostering collaboration and have a broad influence on the department's research efforts.

- **Eigenvector centrality:** This metric considers not only the number of connections a node has, but also the influence of those connections. Nodes connected to other highly connected or influential nodes are assigned a higher eigenvector centrality. This measure identifies researchers whose collaborations extend to key individuals within the network, offering a deeper perspective on leadership and influence in research activities.

- **Betweenness centrality:** This metric measures how often a node acts as a bridge along the shortest paths between other nodes. Faculty members with high betweenness centrality play a critical role in connecting otherwise isolated groups within the network. These individuals facilitate the flow of information and collaboration across different research areas.

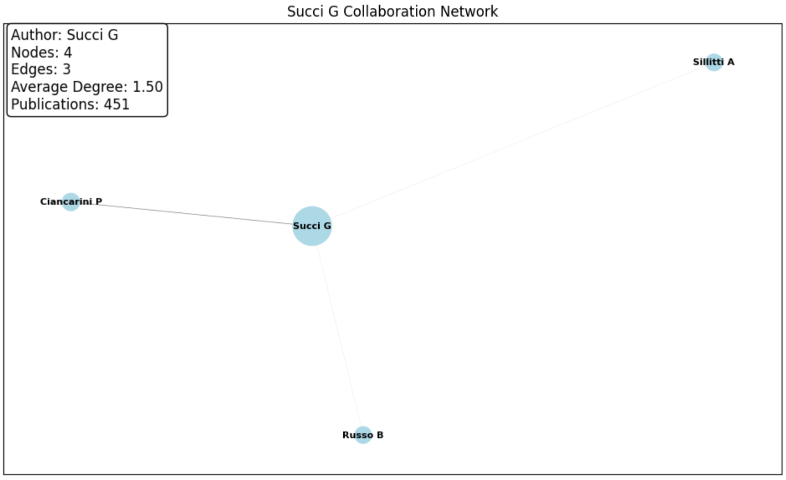
- **Closeness centrality:** Closeness centrality assesses how quickly a node can reach all other nodes in the network. Faculty members with high closeness centrality are well-positioned to access or disseminate information efficiently. Such individuals serve as key points of contact for promoting initiatives or sharing critical information across the department.

- **Local clustering coefficient:** This metric indicates the extent to which a node's neighbors are also connected to each other, reflecting the local cohesiveness of the network. Researchers with high clustering coefficients often belong to denser research groups or clusters. This measure is useful for identifying thematic collaboration groups and understanding areas of intense and specialized research activity.
- **Cliques:** Cliques are subsets of nodes where every node is directly connected to every other node. They represent tightly-knit groups of collaborators focused on specific research topics. By identifying cliques, we can highlight the most cohesive research teams.
- **Community detection (Louvain method):** The Louvain method is used to detect communities within the network by optimizing modularity, a measure of how well the network can be divided into distinct groups. These communities often represent clusters of researchers working on similar topics. Understanding the structure of these communities can reveal hidden patterns, foster collaborations across different groups, and identify opportunities for interdisciplinary research.
- **Density:** Density measures the proportion of actual connections to all possible connections in the network, providing a global view of the network's cohesiveness. A higher density indicates a greater level of collaboration across the department, while lower density may suggest fragmentation or isolated clusters.
- **Resilience:** Examines the network's ability to maintain its connectivity when nodes are removed, highlighting its robustness. This was evaluated through two approaches:
  - *Key node removal:* Simulating targeted attacks by removing nodes with the highest centrality to assess their impact on the network structure. This reveals the network's dependence on highly influential researchers.
  - *Random node removal:* Simulating random failures or disruptions to evaluate the network's vulnerability under non-targeted scenarios.

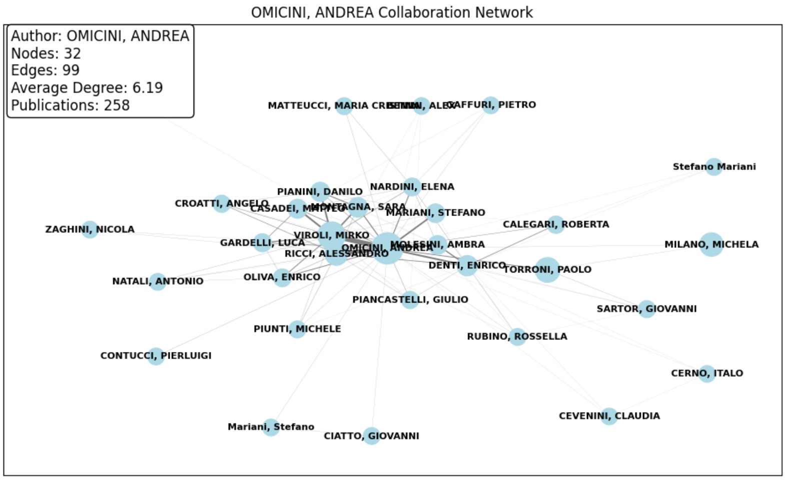
  These evaluations provide insights into critical points in the network and its ability to sustain collaboration despite disruptions. Simulating such attacks also highlights the network's structural resilience and suggests strategies for maintaining connectivity under adverse conditions. By identifying vulnerabilities, the department can better plan for researcher succession and mitigate risks associated with turnover.

5.2 Results

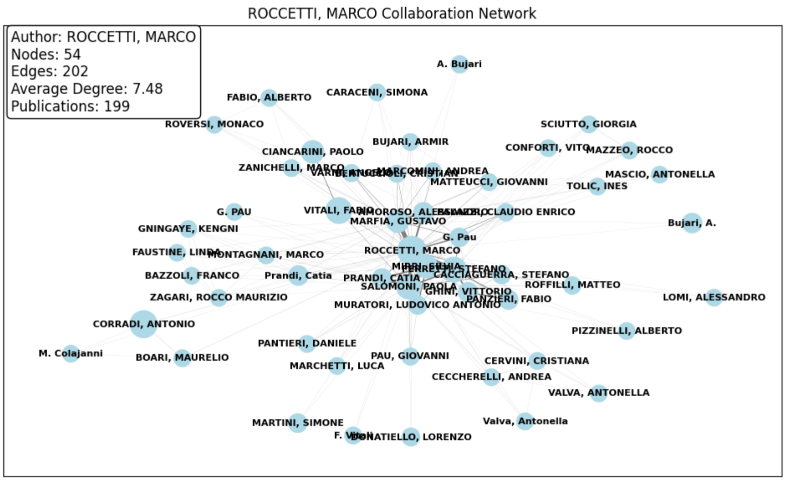
In this section, we present the results obtained from the performed analyses.
To provide a more detailed and focused examination, the analyses were conducted on specific subgraphs centered on the three faculty members with the highest number of publications, as shown in Fig.1. These subgraphs represent the immediate collaboration networks of the selected individuals, offering valuable insights into their roles within the network.
By analyzing these prominent nodes, we examine their influence on the overall network structure, uncover patterns of academic collaboration, and evaluate the dynamics within



(a) Succi G. subgraph



(b) Omicini Andrea subgraph



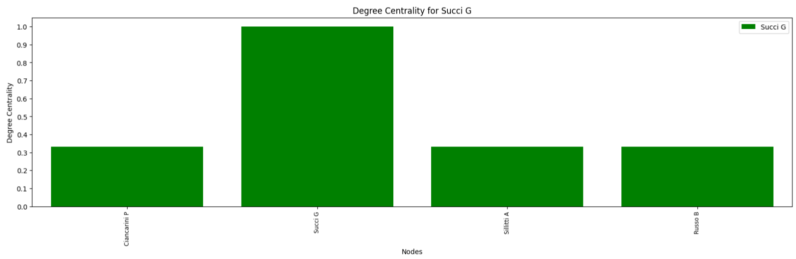
(c) Roccetti Marco subgraph

Figure 1: Faculty members with the highest number of publications.

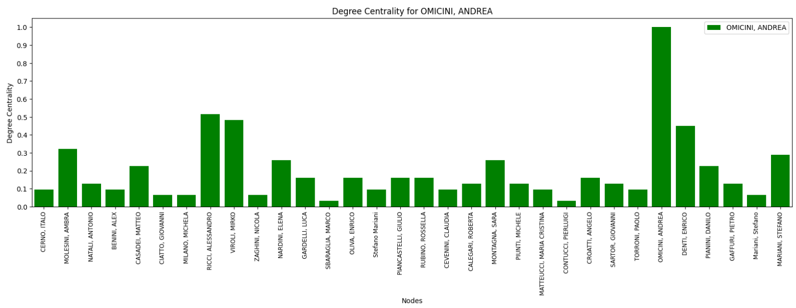
their respective research communities. This approach not only highlights the structural significance of these individuals but also reveals broader trends in research partnerships across the department.

5.2.1 Degree centrality

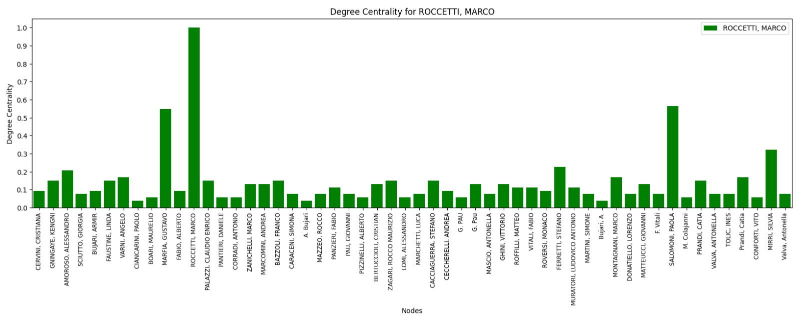
Our analysis of the network began by identifying the most collaborative individuals through degree centrality metrics. By focusing on subgraphs for Professors with the highest publication counts, we highlighted key hubs in the network. The results, presented in Fig. 2, reveal that Succi G., Omicini A., and Roccetti M. exhibit the highest degree centrality scores in their respective subgraphs, highlighting their pivotal roles in fostering academic collaborations.



(a) Succi G.



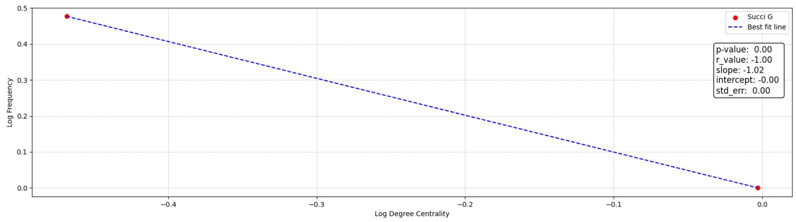
(b) Omicini Andrea



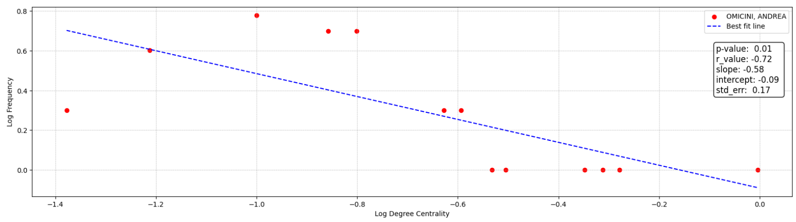
(c) Roccetti Marco

Figure 2: Degree centrality

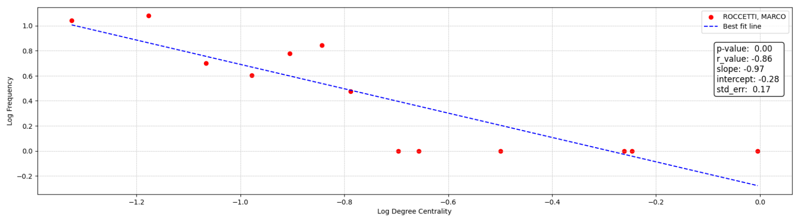
To identify additional key collaborators, a threshold was applied to the degree centrality values. Nodes exceeding this threshold represent the primary collaborators of the top Professor, revealing a tightly connected core of researchers. This analysis emphasizes the hierarchical nature of the network, with the most connected nodes forming a central hub.



(a) Succi G.



(b) Omicini Andrea



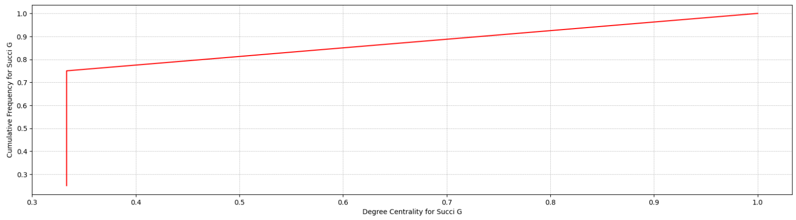
(c) Roccetti Marco

Figure 3: Degree centrality distribution

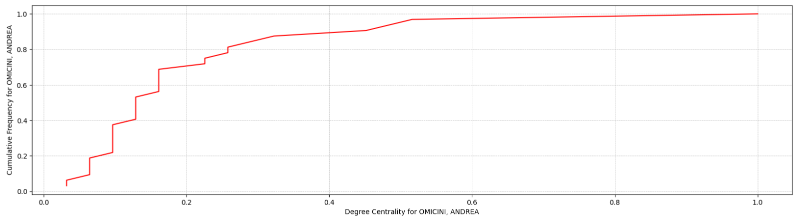
Specifically, each of these Professors surpasses the defined threshold of *0.7*, indicating their prominence as central nodes within their networks. This finding underscores their importance in connecting researchers and coordinating collaborative efforts.
Additionally, we analyzed the degree distribution to understand the structural properties of the network. A log-log plot of the degree centrality distribution, as illustrated in Fig. 3, demonstrates that the network exhibits a power-law behavior, suggesting scale-free properties. The linear regression analysis excluding the top Professor (an outlier due to their exceptional degree centrality) yielded a Pearson correlation coefficient $R \approx -0.87$, with a corresponding p-value < 0.01. These results confirm that the network follows a scale-free distribution, where a few nodes act as hubs with significantly higher centrality compared to others.
Lastly, the cumulative distribution of degree centrality scores was computed, as displayed in Fig. 4. The plot shows a stepped increase in values due to the discrete nature of normalized degree scores. The cumulative analysis further supports the interpretation of a scale-free network, reflecting the heterogeneity in collaborative patterns among faculty members.

5.2.2 Eigenvector centrality

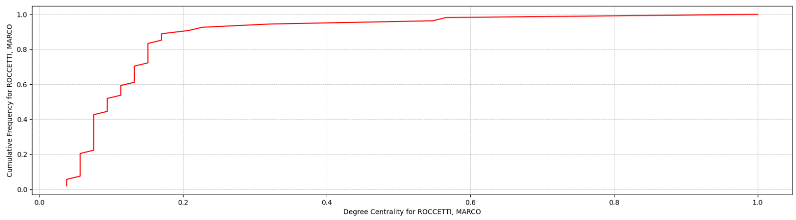
Eigenvector centrality offers a complementary perspective on the importance of nodes, assigning higher values to those connected to other high-centrality nodes. This measure goes beyond degree centrality by emphasizing the quality of connections over mere quan-



(a) Succi G.



(b) Omicini Andrea



(c) Roccetti Marco

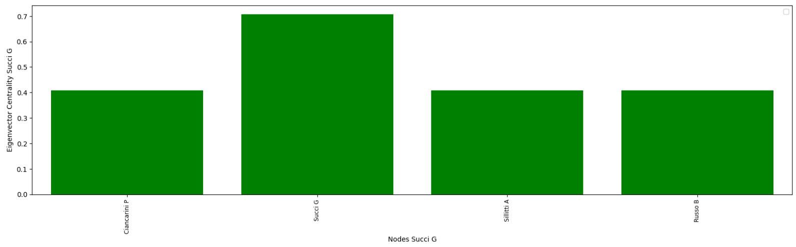
Figure 4: Cumulative distribution for degree centrality

tity. Fig. 5 qualitatively illustrates the eigenvector centrality values for the Professor with the most publications and their network.

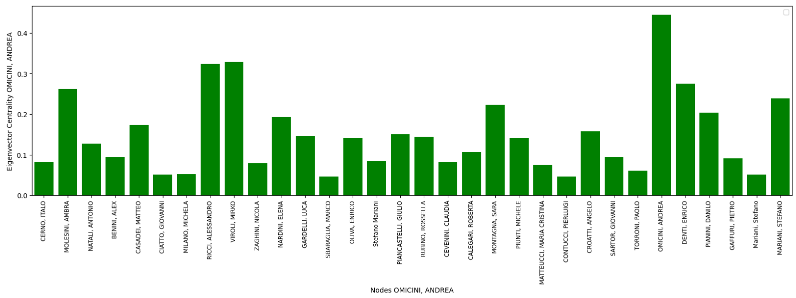
Compared to degree centrality, we observe a different distribution, with some nodes gaining increased significance due to their connection to influential collaborators, while others lose prominence. More in specific Succi G. emerges as one of the highest-ranking nodes with an eigenvector centrality of *0.707*, reflecting the strategic position in a network of influential collaborators, including Russo B. and Sillitti A., who also exhibit high centrality scores (*0.408*).

In contrast, OMICINI ANDREA shows a centrality score of *0.444*, reflecting a strong position in his sub-network, particularly with collaborators like VIROLI MIRKO (*0.328*) and RICCI ALESSANDRO (*0.324*). Notably, a number of other collaborators exhibit relatively lower centrality scores.

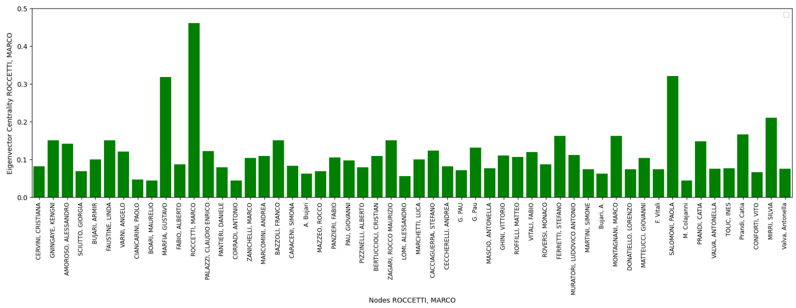
ROCCETTI MARCO demonstrates a similarly high eigenvector centrality of *0.461*, signaling strong connections with influential figures like MARFIA GUSTAVO (*0.318*) and SALOMONI PAOLA (*0.321*), suggesting their central role within their academic community.



(a) Succi G.



(b) Omicini Andrea

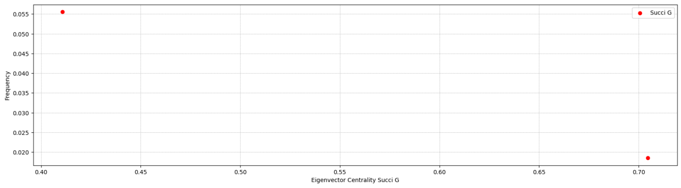


(c) Roccetti Marco

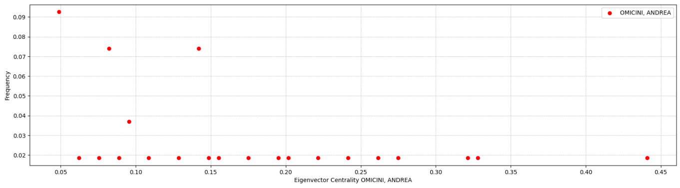
Figure 5: Eigenvector centrality

The threshold analysis reveals that several nodes, including Russo B., Succi G., Sillitti A., and Ciancarini P., surpass an eigenvector centrality threshold of *0.3*, indicating their high relative importance in their networks. This is further underscored by the cumulative distribution of centrality values, where a small number of highly connected Professors significantly contribute to the overall centrality of the network.

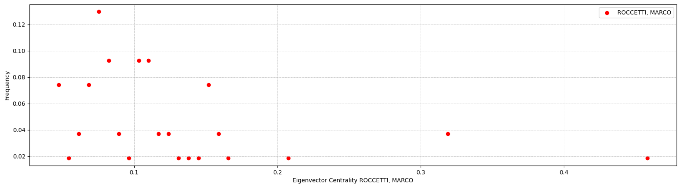
To examine the network's structural properties further, we analyzed the eigenvector centrality distribution. As shown in Fig. 6, the scatter plot of eigenvector centrality values displays a right-skewed distribution, while the cumulative plot in Fig. 7 follows a stepped pattern, as expected from a scale-free network. These results highlight the network's hierarchical structure, where a few nodes with high eigenvector centrality act as bridges to connect less prominent nodes to the network's core.



(a) Succi G.

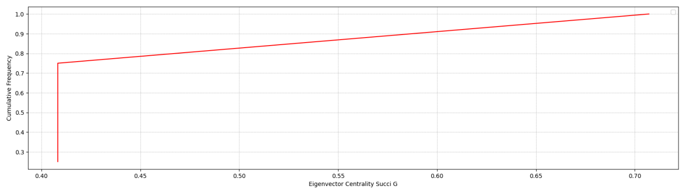


(b) Omicini Andrea

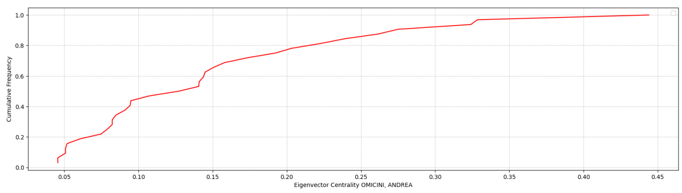


(c) Roccetti Marco

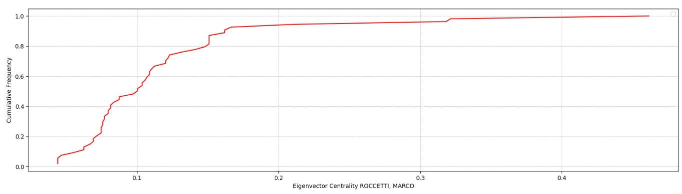
Figure 6: Eigenvector centrality distribution



(a) Succi G.



(b) Omicini Andrea

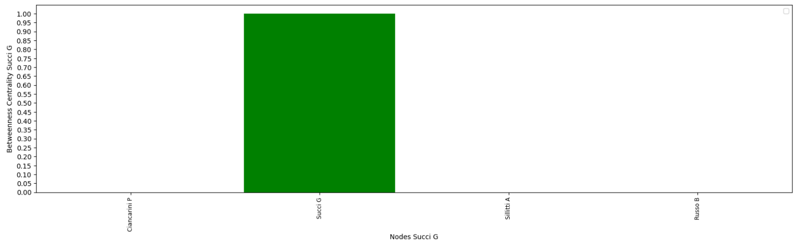


(c) Roccetti Marco

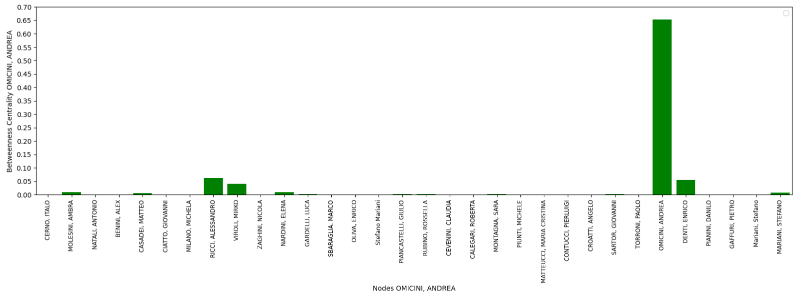
Figure 7: Eigenvector centrality of cumulative distribution

5.2.3 Betweenness centrality

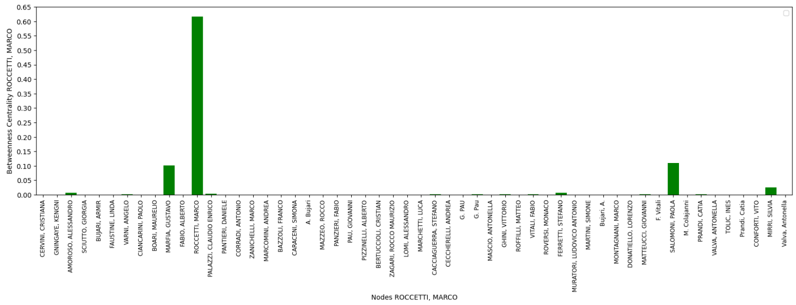
The betweenness centrality values for the nodes considered are shown in Fig. 8. As in the case of degree centrality, most nodes in the network exhibit low betweenness centrality, indicating that they are not frequently part of the shortest paths connecting other nodes.



(a) Succi G.



(b) Omicini Andrea

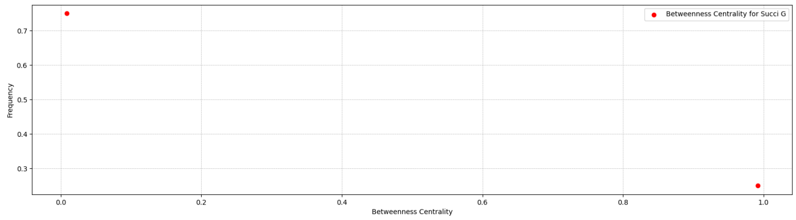


(c) Roccetti Marco

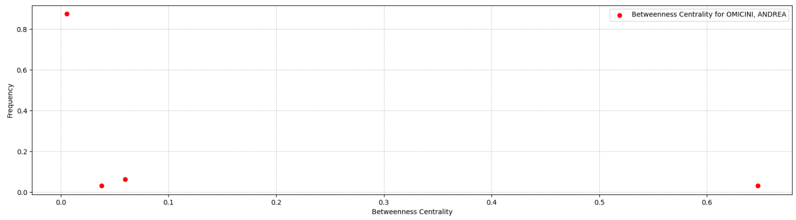
Figure 8: Betweenness centrality

We applied a threshold of *0.01* to highlight nodes with higher betweenness centrality, which were identified as important in connecting other nodes in the network. These nodes are often not the most highly connected in terms of degree or eigenvector centrality, but they play a key role in bridging interactions between other nodes that may not be directly connected. For example, in the subgraph of Succi G., the centrality is concentrated on a small number of key nodes, with Succi G. having the highest value of *1.0*, making him a critical intermediary.

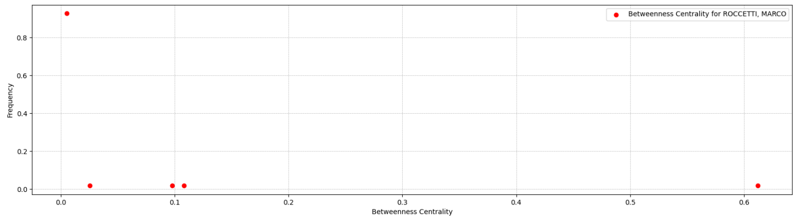
The distribution of betweenness centrality values for the subgraph is shown in Fig. 9. Most of the nodes have values close to zero, with a few nodes exhibiting significantly higher values. This indicates that only a few nodes are acting as important intermediaries in the network.



(a) Succi G.



(b) Omicini Andrea



(c) Roccetti Marco

Figure 9: Betweenness centrality distribution

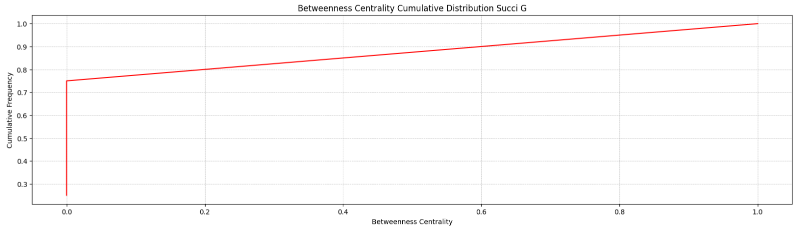
Furthermore, the cumulative distribution of betweenness centrality values is presented in Fig. 10. The plot shows an increasing curve, which is typical for centrality measures in networks, as more nodes reach higher values in relation to the entire network.
Finally, it's important to highlight that while betweenness centrality emphasizes the importance of nodes that serve as intermediaries, it differs from other measures like degree or eigenvector centrality. This is reflected in the different set of nodes identified as important by betweenness centrality, revealing a more nuanced view of the network's structure.

5.2.4 Closeness centrality

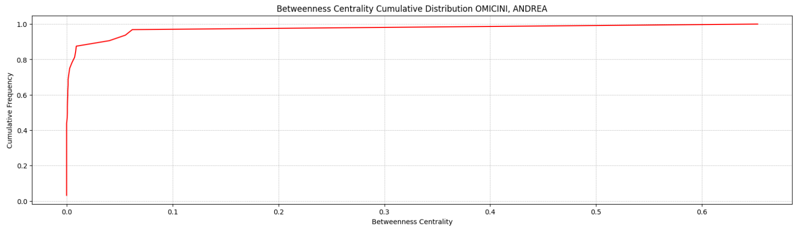
To better understand the positional influence of prominent authors, we calculated the mean closeness centrality values for the three researchers with the most publications:

- **Succi G.**: *0.700*
- **Omicini Andrea**: *0.564*
- **Roccetti Marco**: *0.544*

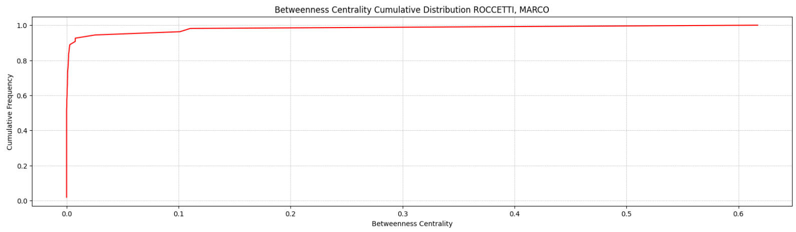
These results indicate that Succi G. maintains the most efficient access to other authors within their collaboration network, likely playing a key role in facilitating interactions and acting as a central figure in bridging academic efforts. Comparatively, Omicini Andrea and Roccetti Marco also exhibit notable but lower efficiency in accessing the network, highlighting their positions as central collaborators within their respective domains.



(a) Succi G.



(b) Omicini Andrea



(c) Roccetti Marco

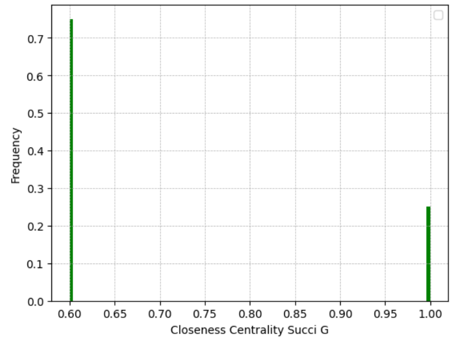
Figure 10: Betweenness Centrality cumulative distribution

To visualize the distribution of closeness centrality values, we generated a histogram (see Fig. 11). This representation reveals that the majority of nodes in the network exhibit lower closeness centrality values, underscoring the prominence of these three authors as key connectors within the broader collaboration network.
Overall, the analysis highlights the differing levels of centrality among the top authors and emphasizes the role of closeness centrality in identifying those who act as efficient communicators within the network.

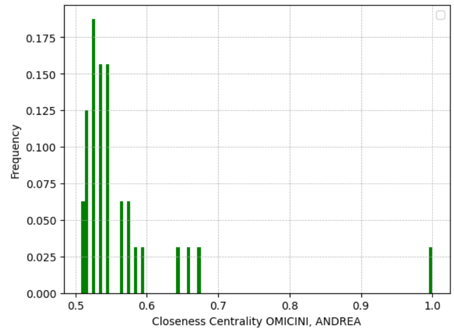
5.2.5 Clustering coefficient

For the authors in this analysis, the local clustering coefficients for each author's network were computed, and the results show the following trends.

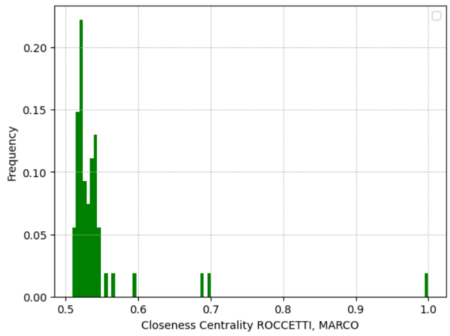
Succi G. The clustering coefficients for the nodes in Succi G.'s collaboration network are consistently low, with all values being *0*. These results suggest that Succi G.'s collaborators tend not to form tightly-knit groups or clusters. Specifically, for Succi G., the values for collaboration with co-authors like Ciancarini P., Sillitti A., and Russo B. are all *0*, indicating no clustering among these collaborators.



(a) Succi G.



(b) Omicini Andrea

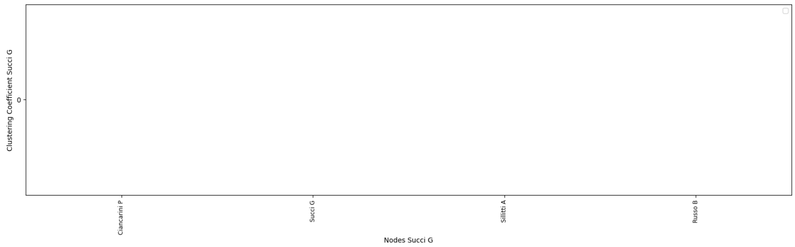


(c) Roccetti Marco

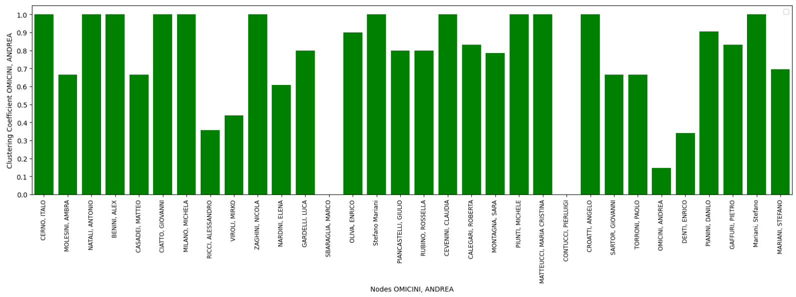
Figure 11: Closeness Centrality

Omicini Andrea Omicini Andrea's collaboration network shows more variation in clustering coefficients, with several values reaching higher levels, such as CERNO ITALO, NATALI ANTONIO, and PIANINI DANILO, all showing clustering coefficients of *1.0*. This indicates that these authors are part of tightly-knit groups where collaborators tend to also collaborate with each other.
In contrast, some collaborations like those with RICCI ALESSANDRO (*0.358*) or DENTI ENRICO (*0.34*) are much lower, suggesting that those particular collaborations are less interconnected with other authors in the network.

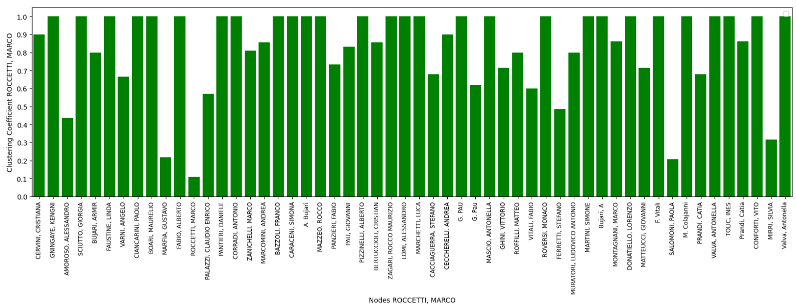
Roccetti Marco Roccetti Marco's network is also characterized by mixed clustering coefficients. Several collaborations have high clustering values, such as GNINGAY KENGNI (*1.0*), SCIUTTO GIORGIA (*1.0*), and CIANCARINI PAOLO (*1.0*), indicating cohesive groups within the network. However, there are also collaborations with low clustering values, such as MARFIA GUSTAVO (*0.219*), reflecting less tightly connected subgroups.
The overall clustering coefficient distributions are visualized in a bar plot, where the variation among authors is clearly seen. As shown in Fig. 12, Succi G.'s network is largely composed of isolated nodes (low clustering coefficients), while Omicini Andrea and Roccetti Marco demonstrate more diverse collaboration patterns, with both high and low clustering values across their networks.
Additionally, the relationship between the local clustering coefficient (C_i) and degree centrality (d) was examined. The scatter plot in Fig. 13 highlights the correlation between these two measures, revealing that authors with higher degree centrality do not necessarily



(a) Succi G.



(b) Omicini Andrea

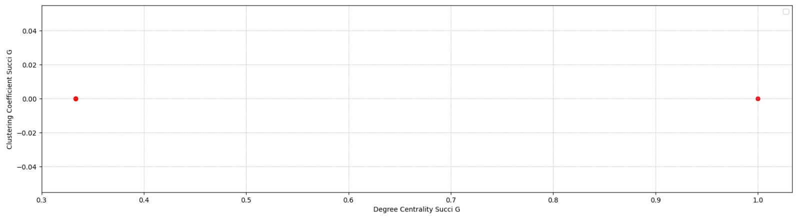


(c) Roccetti Marco

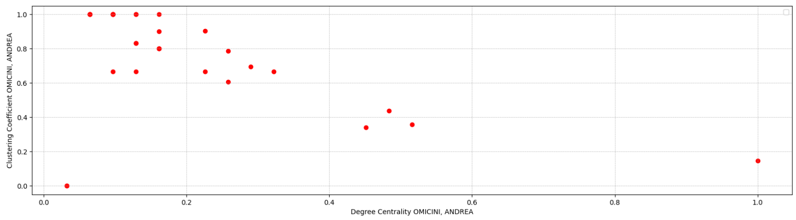
Figure 12: Clustering coefficient

have higher clustering coefficients. For example, Omicini Andrea's co-authors with high degree centrality do not always form tightly-knit collaboration groups, suggesting that some of these authors serve as hubs, connecting a diverse range of collaborators without forming dense clusters.

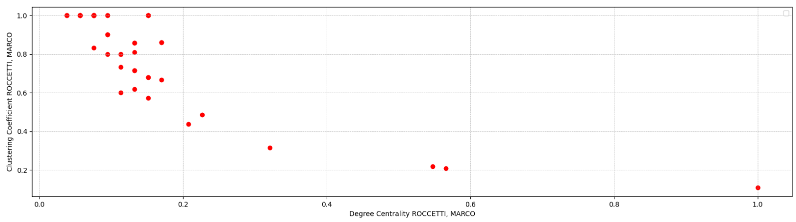
In summary, Succi G.'s network exhibits a more fragmented structure with isolated collaborations, while Omicini Andrea and Roccetti Marco have more connected and cohesive networks with some variation in clustering. The negative correlation between the clustering coefficient and degree centrality suggests that the higher-degree authors tend to have less clustered collaboration groups, potentially reflecting a more diverse range of collaborations that span different research areas.



(a) Succi G.



(b) Omicini Andrea



(c) Roccetti Marco

Figure 13: Closeness Centrality

5.2.6 Cliques

In this analysis, the following cliques were found within the subgraphs of each author.

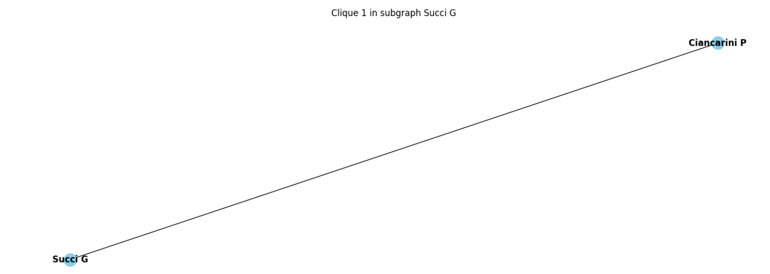
Succi G. The subgraph of Succi G. contains 3 cliques, indicating a relatively low level of tightly-knit collaboration groups. These cliques are smaller and suggest that Succi G.'s collaborations are not concentrated in large, interconnected groups.

Omicini Andrea In contrast, the subgraph of Omicini Andrea reveals 24 cliques, which suggests a more complex network of collaborations. These cliques point to a larger number of smaller groups of authors who are highly interconnected, potentially indicating more frequent or specialized collaboration in research topics.

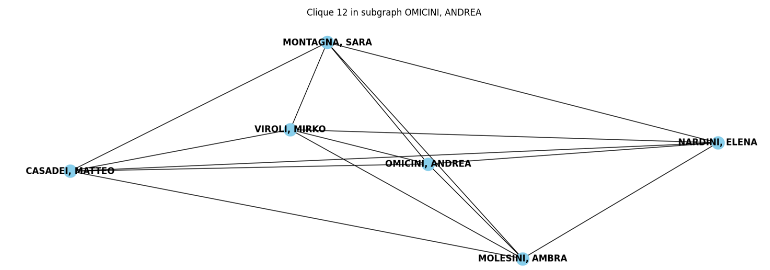
Roccetti Marco The subgraph of Roccetti Marco is characterized by 40 cliques, the largest number of any of the three authors. This result highlights a network with numerous tightly-knit groups, showing that Roccetti Marco is involved in several clusters of collaborators, each representing focused areas of research.

These findings suggest that Omicini Andrea and Roccetti Marco have more fragmented and diverse collaboration networks with several highly connected subgroups, whereas Succi G.'s network is more isolated with fewer interconnected groups.

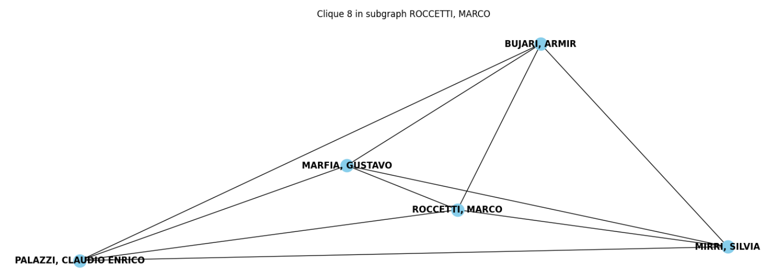
To provide a representation of the collaboration structure within the subgraphs of each author, the plots in Fig.14 show a single example clique for each author. While each author's subgraph contains multiple cliques, these graphs are intended to serve as illustrative examples, highlighting the interconnected nature of these tightly-knit groups without exhaustively depicting all cliques.



(a) Clique 1 Succi G.



(b) Clique 12 Omicini Andrea



(c) Clique 8 Roccetti Marco

Figure 14: Cliques

The clique structure is linked to other network properties, such as degree assortativity, which reflects the tendency of nodes with similar degrees (i.e., number of connections) to connect with each other. In the case of the degree assortativity coefficient, the results for each author are as follows:

Succi G. The degree assortativity coefficient is *-1.0*, indicating a highly disassortative network. This suggests that nodes with high degrees (more connections) tend to connect with nodes having low degrees, while nodes with low degrees tend to connect with those of higher degrees. This is a clear sign that central (high-degree) authors in Succi G.'s network tend to interact with peripheral (low-degree) authors.

Omicini Andrea The degree assortativity coefficient for Omicini Andrea is *-0.3997*, which still indicates a disassortative trend, but less pronounced than for Succi G. It suggests that there is a moderate tendency for high-degree authors to collaborate with lower-degree authors, but not as strongly as in Succi G.'s network.

Roccetti Marco With a degree assortativity coefficient of *-0.3444*, Roccetti Marco's network also exhibits disassortative behavior, though the correlation is weaker compared to both Succi G. and Omicini Andrea. This again suggests that Roccetti Marco's high-degree authors are more likely to collaborate with lower-degree authors, reflecting a more diverse set of collaborations.

These assortativity results suggest that all three authors' networks display a disassortative pattern, where more central authors (those with many collaborators) are more likely to collaborate with peripheral authors (those with fewer collaborators).This could reflect a trend where influential researchers engage with a broader range of researchers, including those who may not be as well connected, potentially fostering cross-disciplinary collaborations or collaborations with emerging researchers.

5.2.7 Density

The density of a network is a measure of the proportion of possible edges that are actually present. In a fully connected network, the density would be *1*, whereas a sparse network would have a density closer to *0*. A higher density indicates a more interconnected network, while a lower density suggests a more fragmented or sparse network structure. For the subgraphs representing the three authors the density values are as follows:

- **Density for Succi G: *0.5***
  The subgraph for Succi G. exhibits a density of *0.5*, which suggests a moderately dense network where about half of the potential edges are actually present.

- **Density for OMICINI, ANDREA: *0.1996***
  The subgraph for OMICINI, ANDREA has a lower density of *0.1996*, which indicates a much sparser network. This suggests that while there are still some significant connections, the network has a larger number of unconnected nodes or relatively few collaborations compared to its possible maximum.

- **Density for ROCCETTI, MARCO: *0.1412***
  The density for ROCCETTI MARCO's subgraph is *0.1412*, which is even lower than the previous two. This further highlights a more fragmented network, where the number of connections is relatively small in comparison to the potential maximum number of edges.

These density values reflect the differences in the collaboration structures of each author's network. Succi G.'s network is more tightly connected, while the networks for OMICINI ANDREA, and ROCCETTI MARCO show more dispersed collaboration patterns.

5.3 Metrics involving the whole network

5.3.1 Community detection

Community detection aims to identify groups of nodes that are more densely connected to each other than to the rest of the network. To achieve this, we used the Louvain algorithm, which partitions the graph by maximizing modularity. Modularity measures the strength of division of a network into communities, helping to identify structures where nodes within a community are more strongly connected than to nodes outside it.
The three largest communities identified are:

- **Community 4**: This is the largest community, consisting of 207 nodes. It represents a group of authors that are densely interconnected, indicating a central and cohesive cluster within the overall collaboration network.
- **Community 14**: The second-largest community consists of 201 nodes. Like Community 4, this group of authors shows a strong internal connectivity, suggesting another distinct and influential collaboration cluster.
- **Community 8**: This community includes 157 nodes, making it the third-largest. Although smaller than the previous two, it still represents a notable subset of the network, characterized by a dense internal structure.

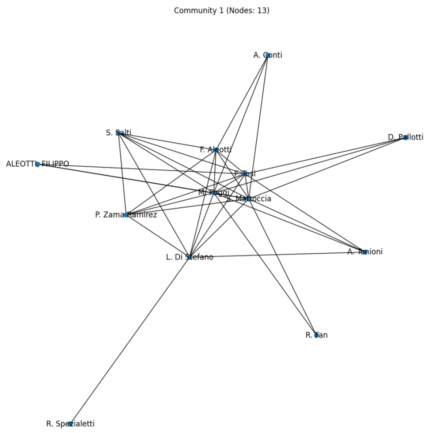
These findings highlight the presence of several strong collaboration groups within the network. The total number of nodes in the graph is 2831, and the network contains 14048 edges, indicating a dense and well-connected collaboration ecosystem.

For better graph readability, we will not display the larger communities described above as examples. Instead, in Fig.15 we will show smaller communities composed of fewer nodes and edges in.

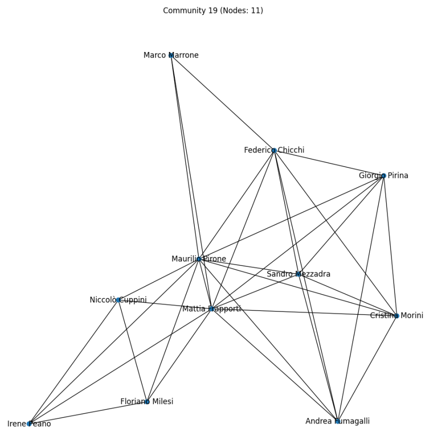
5.3.2 Resilience

The resilience of a network refers to its ability to maintain connectivity when nodes (or edges) are removed. To assess resilience, two common strategies are used: the removal of key nodes and the simulation of random attacks. In this analysis, we tested the resilience of the graph by examining its response to these two scenarios.

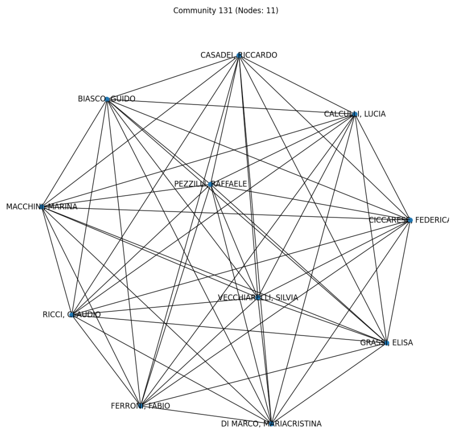
- **Removal of key node**
  Before and after removing a key node, the number of connected components in the graph was *113*. This suggests that network might have redundant paths. Even if a key node is removed, its neighbors could remain connected through other nodes or edges, preserving the structure of the connected components.
- **Simulation of random attack**
  In contrast, when a random attack was simulated, the number of connected components also remained at *113*. This shows that the network's connectivity remained largely unaffected by the random removal of nodes. Random attacks tend to be less damaging to the structure of the network than targeted attacks on key nodes, which suggests that the network is more robust to random disruptions.



(a) Community 1



(b) Community 19



(c) Community 131

Figure 15: Community detection

5.3.3 Assortativity and density

The degree assortativity coefficient of the graph was calculated to be *0.7487*, indicating an assortative structure where nodes with similar degrees tend to be more likely to be connected. This suggests that nodes with high connectivity are more likely to form connections with other highly connected nodes, contributing to the overall network's stability.

The density of the overall network is quite low at *0.0035*, indicating that the graph is sparse, with a small proportion of possible edges actually being present. Despite its sparsity, the network showed resilience in terms of connectivity when subjected to the node removal and random attack simulations. As said, the subgraphs for Succi G, OMICINI ANDREA, and ROCCETTI MARCO exhibit higher density values (e.g., Succi G at *0.5*), reflecting more interconnected collaboration patterns within each individual author network.

These results highlight the network's ability to withstand random disruptions but also show potential vulnerabilities when critical nodes are removed. The resilience of the network is influenced by its degree assortativity and overall density, which play a role in how connected the network remains after node removals.

6 Conclusion

This study provides an analysis of the collaboration networks of three authors: Succi G., Omicini Andrea, and Roccetti Marco. By exploring various network properties such as cliques, community detection, density, and resilience, we gained valuable insights into the structure and dynamics of their respective collaboration networks.

The analysis of cliques revealed distinct differences in the collaboration patterns of the authors. Succi G.'s network is characterized by fewer, smaller cliques, indicating a relatively isolated set of collaborations. In contrast, Omicini Andrea and Roccetti Marco exhibit more fragmented and diverse networks, with a larger number of tightly-knit subgroups. This suggests that both Omicini and Roccetti are involved in more specialized or varied research areas, where collaborations are often concentrated within smaller, more focused groups of authors.

Community detection further highlighted the presence of strong collaborative subgroups within the overall network. The Louvain algorithm identified three significant communities, each representing a dense and interconnected group of authors. These findings align with the clique analysis, where Omicini and Roccetti's networks are more fragmented into numerous smaller, influential communities, while Succi G.'s collaborations appear more centralized.

When examining density, we observed notable differences between the authors' networks. Succi G.'s subgraph stands out with a much higher density, reflecting a well-connected network with a significant number of collaborations. The networks of Omicini and Roccetti are more sparse, with relatively lower density values, indicating that their collaborations are more dispersed and less interconnected. This highlights a more fragmented approach to collaboration, where authors may work across various, less cohesive research clusters.

In terms of resilience, the network demonstrated exceptional robustness to disruptions. The connectivity remained stable even after the removal of both randomly selected and key nodes, indicating a strong resistance to fragmentation. This high resilience can be attributed to the network's degree assortativity and overall sparsity, which ensure that nodes with similar degrees are more likely to maintain connections, enhancing the network's cohesion and stability against potential disruptions.

Together, these findings reveal the complex nature of academic collaboration networks. While Succi G.'s network appears more centralized and cohesive, Omicini and Roccetti's networks are more decentralized and fragmented, with numerous specialized collaboration clusters. The study wants to underscore the importance of network structure in understanding academic collaboration patterns, as well as the resilience of these networks under different conditions.

7 Critique

This study provides a detailed quantitative analysis of the collaboration networks of Succi G., Omicini Andrea, and Roccetti Marco, addressing the research problem of understanding the structural dynamics within academic collaboration networks. The approach of analyzing cliques, community detection, network density, and resilience offers valuable insights into the connectivity and robustness of these networks, thus partially solving the problem presented.

The findings provide a clear view of the differences in collaboration patterns across the three authors, illustrating varying levels of fragmentation and connectivity. However, the analysis could be considered as only a partial solution to the problem, as it primarily focuses on structural aspects of the networks without delving into deeper insights regarding the specific factors driving these patterns, such as the influence of particular research topics, funding sources, or external collaborations.

In terms of what could have been done further more, in first place, incorporating additional data, such as the nature of collaborations (e.g., co-authored publications, project participation), would have provided a more nuanced understanding of the collaboration dynamics; then, a more comprehensive model could also include time-based analyses to track changes in collaboration patterns over time and assess how these patterns evolve. Lastly, a broader dataset encompassing more authors or institutions would have offered a more holistic perspective on academic collaboration networks, providing a clearer comparison between different academic domains or research disciplines.

In conclusion, while this study addresses key aspects of the collaboration network problem, there are several opportunities for deeper exploration and refinement. Expanding the dataset, applying alternative methodologies, and considering additional variables could have provided a more comprehensive answer to the research question, enhancing the study's scope and depth.

References

[1] NetworkX Developers. Networkx: Network analysis in python. https://networkx.org/.

[2] Department of Computer Science and University of Bologna Engineering. Disi publications. https://disi.unibo.it/it/ricerca/pubblicazioni?&pagesize=7523.

[3] The Pandas Development Team. Pandas: Python data analysis library. https://pandas.pydata.org/.

[4] Lucrezia Ilvento Valerio Desiati, Erik Koci. Sna project. https://github.com/valeriodesiati/SNA-project, 2024.