Influence of Multi-dimensional Social Capital on Structure of Scientific Collaboration Networks: *MDSC@SciCoNet*

Avijit Gayen[1,2*], Somyajit Chakraborty[3], Nilotpal Chakraborty[1] and Angshuman Jana[1]

[1]Department of CSE, Indian Institute of Information Technology Guwahati, Bongora, Guwahati , 781015, Assam, India.
[2*]Department of CSE, Techno India University, West Bengal, EM-4, Sector V, Salt Lake, Kolkata, 700091, West Bengal, India.
[3]Department of CSE, University College Cork, Cork, Ireland.

*Corresponding author(s). E-mail(s): avijit.gayen@iiitg.ac.in; Contributing authors: somyajitChakraborty@ucc.ie; nilotpal@ieee.org; angshuman@iiitg.ac.in;

Abstract

Understanding the existence and effects of social capital in various social systems has been a focus for social scientists for many years. Research indicates that social capital is a social attribute influenced by the structural, relational, and cognitive components of an individual's social network. Social capital can significantly impact career success, and its effects have been studied in various domains, including scientific collaborations. In scientific collaborations, an author's social capital has been shown to influence the scientific impact of their research. However, few studies have explored how components of social capital contribute to variations in collaboration network structures across different scientific fields. In this paper, we analyze the evolution of scientific collaboration networks across different domains by examining the social capital components of authors within the network. For structural and relational capital measures, we use degree centrality and the count of prolific coauthors, respectively. Given the importance of tie strength in both relational and cognitive capital, we separately investigate the role of tie strength among collaborators in the evolution of these networks across various domains. We specifically examine the attachment probabilities of nodes in the network based on: a) degree centrality, b) count of prolific coauthors, and c) tie strength among collaborators. We then model these attachment probabilities and validate them with empirical data. Our observations reveal that while the attachment of nodes with respect to these parameters is preferential, their impact on network structure varies. This study identifies key differences in collaboration dynamics between CSE and physics. In CSE, new collaborations are driven by network degree and prolific co-authors, with minimal influence from social tie strength. In contrast, physics shows a stronger role for social ties in collaboration formation. Both fields emphasize node tie strength in republishing, with physics favoring smaller, closed groups, while CSE exhibits broader networks. These findings highlight the varying roles of structural and social capital across disciplines.

Keywords: Scientific collaboration networks, Co-authorship networks, Social Capital, Tie Strength, Preferential Attachment, ANOVA

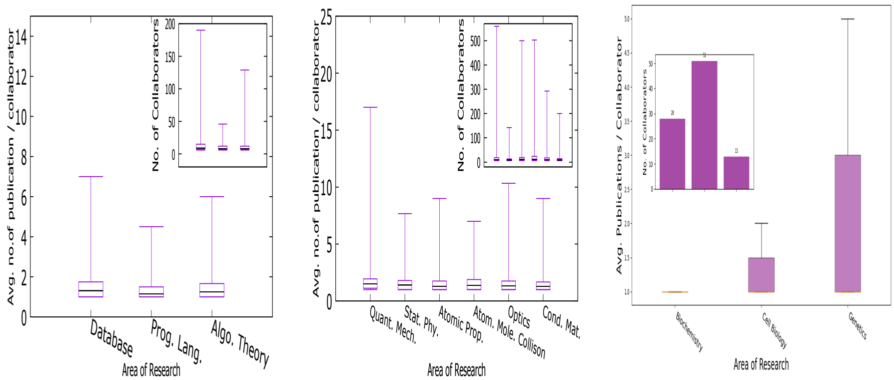
1 Introduction

The existence and impact of social capital across multiple disciplines [1, 2] has long been a focal point of research for scholars. Understanding how social networks, relationships, and collaborative ties influence knowledge creation and innovation has been particularly important in this context. Research indicates that social capital significantly influences the scientific impact of research works [3], and it also promotes new collaborations, leading to innovations and the spread of knowledge within the scientific community [4]. Social capital is rooted in individual social connections within the social structure [5]. The social ties among coauthors lead to the formation of working groups, which in turn determine the social capital of the authors. This social capital can further facilitate new collaborative explorations [6].

However, the strength of the ties among coauthors can indicate their inclination to collaborate more frequently with each other. This affinity is often influenced by factors such as author preference [7] and geographical proximity. Therefore, the interaction of social capital and such social affinity significantly shapes the structure of scientific collaboration networks [8], warranting thorough investigation. In a scientific collaboration network, authors are depicted as 'nodes,' and an 'edge' represents the co-authorship between two authors. A joint publication by two coauthors signifies their collaborative relationship, and the number of their joint publications determines the weight of the edge between them.

Preliminary observations from experiments using our bibliographic dataset, covering specific subfields in Computer Science and Engineering (CSE), Physics and Biology, reveal significant variations in the number of collaborations per author and the number of publications per collaboration (pairwise) across these disciplines (see figure 1). In figure 1, each quartile box represents the minimum, first quartile, median, third quartile, and maximum values of the average number of publications per collaborator. This suggests differences in collaboration patterns and research outputs among these disciplines. Similar findings have been reported in [9], indicating that in some fields, researchers prefer to collaborate within a close-knit group, highlighting the importance of social ties. Additionally, co-authorship has been linked to various factors such as node preferentiality [7], geographical proximity [10], and advancements in communication technologies [11]. However, these studies do not explicitly examine the role of social capital in collaboration.

This highlights the need to understand whether authors' social capital and their connections with coauthors can explain the variations in collaboration networks and research output across different research domains. We believe that investigating this could help predict the growth of a scientific field by analyzing the social capital and ties of its researchers. The formal research question for this study is: *RQ: How can the various dimensions of social capital, along with social ties, explain the topological differences among the various domains of the scientific collaboration network?*



(a) Computer Sc. & Engg. (b) Physics (c) Biology

Fig. 1: It represents the quartile range of the average number of publications per collaborator across different subfields in CSE, Physics and Biology in figure 1(a), 1(b) and 1(c) respectively. In the inset, we plot the number of collaborators for the same authors in those corresponding sub-domains of research.

An author's social capital comprises three components—structural, relational, and cognitive capital [12]—which can be derived from their social neighborhood. Structural capital refers to the researcher's prominence within the research community, often determined by social connectivity measures such as degree centrality, closeness centrality, and betweenness centrality. Relational capital denotes the social importance of an author within the network, typically measured by the number of esteemed authors in their social circle. Cognitive capital assesses an individual's ability to solve problems while working in a team, often measured by factors such as work experience tenure and team exploration [13]. Therefore, cognitive social capital pertains to an individual's problem-solving capability, while structural and relational capital depend on the social connectivity of the nodes. Consequently, the study of structural and relational capital becomes more significant.

In this paper, we first examine the role of structural and relational capital, along with the strength of coauthor ties, in collaborations across various research domains. We then analyze the impact of each social capital component and tie strength in two types of link formation dynamics: (a) the creation of a new link between two authors, and (b) the increase in the weight of an existing link between two collaborators. Our goal is to determine if the relative dominance of any of these parameters can explain the differences in collaboration network topologies across various research fields. Using empirical bibliographic data from the CSE, Physics and Biology disciplines, we investigate whether the dominance of these factors in the collaboration process can account for the variation in collaboration networks across different domains.

Our approach involves empirically observing the attachment probability of authors through coauthorship based on various social capital measures and the strength of their ties with coauthors. We measure the authors' structural capital using parameters like degree centrality, betweenness centrality, and closeness centrality within the collaboration network. Relational capital is determined by the number of prolific coauthors an author has. While we do not explicitly consider team exploration to investigate the role of individual author capabilities in the growth of the collaboration network, an author's publication tenure is implicitly considered through their publication count. Tie strength among coauthors is measured by the number of joint publications. We examine the impact of each of these factors—social capital components and tie strength—on the structure of collaboration networks in three major disciplines, CSE, Physics and Biology. Additionally, we observe for each discipline whether one factor (tie strength or social capital) predominates in the formation of collaborations and the resulting effects.

Our Contributions: The contributions of this paper can be summarized as follows:

1. We examine the role of both structural and relational capital in the attachment of nodes during network evolution, deriving expressions that model the relationship between the attachment probability of the nodes and their respective social capital measures.
2. We also explore the role of tie strength in the attachment of nodes during network evolution, deriving expressions that model the relationship between the attachment probability of the nodes and their tie strength.
3. We examine the role of various components of social capital and social ties in two distinct types of link formation dynamics: (a) the creation of a new

link between two authors, and (*b*) the increase in the weight of an existing link between two collaborators.
4. We analyze the comparative influence of social capital components and tie strength in the publication process across diverse domains, such as CSE and Physics, aiming to elucidate the structural disparities in collaboration networks within these domains.

Although our investigations confirm preferential attachment of nodes based on their structural and relational capital, aligning with established theories on preferential attachment, we interestingly found that the varying dominance of these three factors across domains might explain topological differences. We observe that structural capital, relational capital, and tie strength among coauthors contribute to two prevalent evolution dynamics: a) the formation of new links between nodes in the network, and b) the publication of new works by existing collaborators (which we term as republishing dynamics). Experimental results show that new link formation dynamics dominate over republishing dynamics in the CSE domain, whereas the reverse is more significant in Physics. This finding could account for the topological disparities across research fields.

We have organized the paper as follows: We discuss the detailed related literature in the next section. In section 3, we define the various measures of social capital components and social tie strength in the context of scientific collaboration networks. In this section we also provide the details of the dataset used for our experiments. It also includes the highlights the role of social capital and the effects of social ties among the collaborators in the growth of the network. Subsequently, in section 4, we set up several experiments to observe the role of social capital components and social ties among the coauthors in various types of link formation dynamics in collaboration networks. In section 5, we describe the results of the executed experiments and its inference in this section. Further, we show the results of the experiment to reveal the relative dominance of social capital and social ties across the domains of research on two specific domains of research, i.e., CSE and Physics, to explain the topological growth difference in different domains. We finally conclude and discuss the future scope of research in section 6.

2 Related Works

In this section, we provide an overview of the existing works in the area of the evolution of scientific networks and also highlight the existing literature on social capital and its different components, i.e., structural, relational, cognitive, etc. We outline the significant works that propose measures of social capital in collaboration networks. We subsequently overview the works that deal with social ties among co-authors. For these works, we outline their limitations to motivate the scope of the current paper.

The rigorous study of scientific collaboration networks explored the details of collaboration patterns over the last two decades [9]. These works deal mostly with the nature of scientific networks based on the co-authorship and citation network dynamics. In certain other works, the authors observe the evolution process of international scientific collaborations [14] and international patent network [15]. In contrast, in works like [16], the authors explore the key features of collaboration networks like the number of coauthors in a scientific paper, the distance between the coauthors in the network, and the number of papers of the authors. In some recent work [17], Ribeiro et al. investigate the development and stability of scientific international collaboration. They evaluate the properties of those networks and their long-term behavior. In another recent work [18], Singh et al. discussed the changing collaboration pattern of authors between Indian physicists through various network measures. Although these works provide insights into collaboration networks based on empirical observations, however, they do not provide any mathematical model that could explain the growth process.

Analyzing the growth of co-authorship networks has also been considered in several important works [10]. In [19], Jeong, et al. observed preferential attachment in network evolution of co-authorship networks and also in several diverse fields of interest, e.g. science co-authorship network, internet, actor collaboration network, etc. In [20], Redner et al. observed how citations evolve with time. They show a positive correlation between the citation count of the papers and their average age. Further, in [11], Agrawal et al. analyzed the evolution of scientific collaboration networks with the advancement of communication technology. In [21], Lee et al. have observed that at different stages of time, network evolution follows different patterns, and external influences mostly govern the switching from one pattern to another. They highlighted the influence of large-scale international conferences on scientific collaboration networks.

Some of the existing works attempt to derive analytical models that explain the evolution of scientific collaboration networks [22, 23]. In [22] and [24], the authors considered two different types of dynamics that lead to the network evolution: 1) addition of new nodes in the network and 2) attachment of existing nodes in the network. They observed that preferential attachment based on node degrees plays a crucial role in network evolution. In [25] and [26], the authors proposed an evolutionary model for growth in weighted networks. Their model observed that *"strength"* of a node plays an important role in the network evolution process. In another work [27], the authors proposed a growth model of a scientific collaboration network based on the inter-tuning of three factors — node degree, weight, and a path-dependent preferential attachment mechanism. Recently in [28], Xu et al. proposed a growth model of collaboration networks controlled by factors like preferential and random attachment and network growth speed. In another recent work [29], authors applied a statistical technique to estimate the non-parametric transitivity and preferential attachment simultaneously in a growing scientific collaboration network.

In a recent work [30], Leifeld observed the polarization of academic collaboration in the domain of social sciences in two distinct camps, i.e., hermeneutic and nomological researchers. The author analyzed the entire co-authorship networks in a social science discipline in two separate countries over five years using an exponential random graph model. The empirical analysis reveals that assortative mixing in publication strategies is present and leads to polarization in scientific collaboration. However, there exists some endogenous local factors and their interplay, which reflects microscopic behavior in collaboration development, which affects macroscopic polarization dynamics. This work also observes the topological variation across the different domains of scientific research. In another work, [31], Zhao et al. discussed the growth of the network, which has been analyzed and observed from the different levels, e.g. macro-level and micro-level network measures across the periods. However, none of the models successfully explain the structural difference of the scientific networks across different domains [32]. In [32], the authors represented the collaborations in the different domains of research as a multiplex network and showed that communities coexist and overlap at the different layers (domains); however, they did not attempt to explain the structural variation of the topologies across the domains. In a related study [33], the authors investigate similarities in collaboration patterns not only among researchers but also among R & D firms across various domains and their sub-domains. They find that while these collaboration patterns exhibit certain commonalities, there are notable topological differences between the domains. Their research goes further by proposing an agent-based model designed to simulate the growth of collaboration networks across different domains. This model provides valuable insights into the various factors influencing network growth. However, it is important to highlight that their work does not fully address the role of different dimensions of social capital in the evolution of these networks, which could help to explain the observed topological differences across domains.

The concept of social capital has received good attention in a wide range of social science disciplines (e.g., political science, economics, and organization science) [34, 35]. Works in [36, 37] have shown that the social capital of a system influences the innovation and knowledge diffusion process among the users. In [38], the author defined social capital as the features of social life, networks, norms, and trust - that enable participants to act together more effectively to pursue shared objectives. Several other definitions and measures of social capital have been proposed in the literature [39, 40]. Instead of identifying social capital as a single entity, Tsai et al. in [13] identified three components, i.e., structural, relational, and cognitive capital. The existence of social capital and its impact on scientific collaboration networks has been well-studied in [3, 12]. In [12], authors define measures of the components of social capital and show their influence on research productivity measured in terms of citation count. In a similar kind of work [3], Abbasi et al. have shown the association between scholars' social capital measures and citation-based performance measures. A plethora of works investigates the role of structural capital on the evolution of collaboration networks [41]. Though there exists some work [42] which highlights that the interplay of cognitive and relational social capital dimensions plays a major role in university-industry collaboration, it is yet to be fully understood how various dimensions of social capital of the authors combined with their strength of social ties influence such author-author collaboration network evolution.

As social ties are an important component in network growth, their role in network formation has been studied extensively for various networks [43, 44]. In literature [45] importance of tie in inter-firm collaboration network has been explained as an "instrument for business activity". In [45], the author observed that strong ties are not only used extensively to provide knowledge and information but also to maintain, extend, and enhance business and personal reputations. In another work [43], Hahn et al. have observed the attachment probability of a new developer to a project is highly associated with strong collaborative ties among the existing developer in the context of open-source software development (OSSD). In an earlier work [46], Cross et al. observed that individual performance in knowledge-intensive work is associated with properties of both networks and ties. A large body of works [47, 48], reveals the role of social ties among collaborators in the creation of knowledge, innovation, and scientific problem-solving.

While existing literature extensively examines scientific collaboration networks, focusing on co-authorship, citation dynamics, and network evolution, much of this research addresses individual factors such as preferential attachment, node strength, and external influences like conferences. Although social capital's structural, relational, and cognitive dimensions are recognized for promoting innovation and knowledge diffusion, these components are often studied in isolation, with limited attention to their combined impact. Moreover, recent studies highlight the polarization of academic networks and macro- and micro-level network patterns, but they fall short of explaining how multi-dimensional social capital influences topological differences across research domains. A significant gap remains in understanding how the collective impact of these social capital dimensions shapes collaboration network structures across different fields. In this paper, we address this gap by demonstrating that scientific domain collaborations are driven by researchers' social capital and by explaining how varying influences of social capital components contribute to structural variations across domains. Furthermore, existing research lacks insight into how strong affinities among co-authors affect network structure. While such affinities can form small working groups, it is unclear whether these trends are neutralized when social capital components play a more dominant role in collaboration. This underlines the need for further investigation into the nuanced effects of social capital on network evolution across disciplines.

In the next section, we discuss the proposed work which includes the measures of social capital components, dataset description and empirical study of social capital components as well as the preliminary models of attachment of nodes with respect to the social capital components.

3 Proposed Work

In this section, we outline the components of social capital and their corresponding measures. We then define node tie strength, which gauges the strength of social relationships among researchers. Additionally, we provide a detailed description of the dataset used for the empirical study of the evolution of scientific collaboration networks across various domains and subdomains. We also propose preliminary models for node attachment in the network, based on various components and social ties of scientific collaboration networks.

3.1 Matrices of social capital components

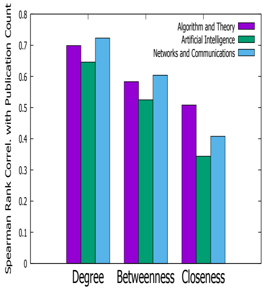
As stated earlier, social capital in scientific collaboration networks has been measured mainly using three different components — structural, relational, and cognitive capital. Structural capital measures the topological importance of an author in the concerned research community, whereas relational capital estimates the importance of the social relations among the collaborators. In literature, cognitive capital has been explained as the cognitive skills acquired over the period of research to solve scientific problems.

The existing works [49, 50] suggest degree, betweenness, and closeness centrality as well-known measures of structural capital. Hence, we adopt a similar methodology to measure structural capital in scientific collaboration networks. Since we mainly focus on collaboration networks, the publications of the authors are the manifestation of collaborations among the authors. Thus, we calculate the correlation between publication count and various centrality measures. We discover through experimental investigations that degree centrality has the highest correlation with publication count across the different domains and subdomains of CSE, Physics and Biology (shown in figure 2). Thus for further analysis, we consider degree centrality as a measure of structural capital.

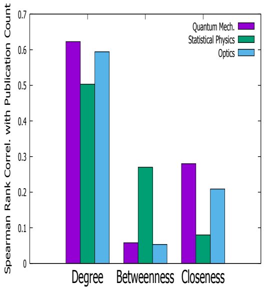
Inspired by the method of measuring relational capital in [12], we also measure the relational capital of an author based on the prolific coauthors with whom she has jointly published. We label an author as prolific when she has published at least three papers during the last four years. In the next 4 years time window, we develop the network and count the prolific coauthors of an author from the earlier labeled author list. Since prior works [51] reveal significant positive effects of cognitive capital on relational capital. In a work [42], Steinmo et al. highlighted that there is a strong interaction between cognitive capital and relational capital in the development of scientific collaborations among academic institutions and industry. Thus we do not explicitly measure cognitive capital for our investigation.

3.2 Matrices of social ties

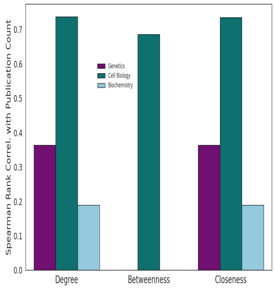
In this section, we initially quantify the link strength, which measures the relational strength between the pair of collaborators. We also introduce the concept of *node tie strength* that provides a measure of the tendency of the



(a) CSE



(b) Physics



(c) Biology

Fig. 2: It represents the Spearman Rank correlation coefficient of publication count with the degree, betweenness, and closeness centrality respectively across different subfields in CSE, Physics and Biology in figure 2(a), 2(b) and 2(c) respectively.

nodes to publish more with the existing collaborators rather than exploring new ones.

Link strength:

Link strength is measured by the normalized value of the joint publication count of two collaborators as a measure of tie strength. If w_{ab} represents the number of joint publication of authors a and b and w_a is the total publications of author a, then the tie strength of edge $a \rightarrow b$ in the collaboration network is denoted by τ_{ab} is represented as:

$$\tau_{ab} = \frac{w_{ab}}{w_a} \tag{1}$$

τ_{ab} describes the collaboration strength of an author with her co-authors. Specifically, collaboration strength signifies the relative importance of the specific co-author among her all co-authors.

Node tie strength:

As stated earlier, node tie strength measures the tendency of the authors to republish with existing collaborators. We measure the node tie strength of a node by the maximum value of the link strengths with all its neighbors. Thus for a node a with neighbor set $N(a)$, the node tie strength is represented as,

$$\tau_a = \max_{i \in N(a)} \{\tau_{ai}\} \tag{2}$$

It essentially measures the reattachment strength of an author with its existing links. A high value of τ_a implies that author a publishes mostly with a small set of coauthors and hence has a lower tendency to explore new collaborations.

We next investigate the importance of these parameters in the formation of new collaborations.

3.3 Dataset Description

We use bibliographical data of publications in the areas of CSE and Physics. We use the dataset in [52] for publications in the areas of CSE. Based on the occurrence of the relevant domain name in the bibliographic records (as shown table 1; the metadata provided includes the following fields: title, authors, affiliations of the authors (continent name), year of publication, topic, the field of research, and index number of the publication), the publications are filtered and classified into “Databases”, “Programming languages,” “Algorithms and theory,” “Artificial Intelligence” and “Networks & Communications”. This produces a separate collaboration network for each classified area of interest. For each of these areas, a weighted collaboration network is formed of the authors where the link weight is measured by the number of publications in that specific domain of research with the corresponding coauthor over a defined time duration. The availability of large data sets in those areas of research helps us to obtain a realistic view of growth dynamics in scientific collaboration networks. We identified about $57,380$, $14,004$, $96,280$, $107,355$ and $54,991$ publications of “Databases”, “Programming languages”, “Algorithms & theory”, “Artificial Intelligence” and “Networks & Communications” respectively that were authored by $60,190$, $14,903$, $58,814$, $109,441$ and $69,240$ numbers of distinct authors.

Table 1: Snippet of dataset (publication details) [52]

#*Transaction Management in Multidatabase Systems.
#@Yuri Breitbart,Hector Garcia-Molina,Abraham Silberschatz
#l,north america,north america,north america
#t1995
#cModern Database Systems
#fdatabases
#index1

The Physics data set is obtained from American Physics Society [53]. It can be accessed online by requesting [1]. The American Physical Society (APS) has made available to researchers two datasets based on their journals, Physical Review Letters, Physical Review, and Reviews of Modern Physics. The dataset comprises 450,000 articles and dates back to 1893. We have used the Article Metadata dataset for this study, which consists of the basic metadata of all APS journal articles. The metadata provided includes the following fields: DOI, journal, volume, issue, first page, and last page OR article ID and the number of pages, title, authors, affiliations, publication history, PACS codes, table of contents heading, article type, and copyright information. We have identified

[1] http://journals.aps.org/datasets

Table 2: General information of raw and filtered APS dataset.

APS dataset	Raw	Filtered
Number of valid papers	1,58,389	1,57,721
Number of papers with no author	668	-
Number of papers with no publication year	0	-
Number of authors	-	64,832
Avg. number of papers per author	-	2.887
Avg. number of authors per paper	-	3.262
Number of coauthorship edges	-	2,40,448

Table 3: General information of raw and filtered PubMed dataset.

PubMed dataset	Raw	Filtered
Number of valid papers	9,999	-
Number of papers with no author	60	-
Number of papers with no publication year	0	-
Number of authors	-	9,868
Avg. number of papers per author	-	1.003
Avg. number of authors per paper	-	7.101

six broad subdomains under Physics e.g. “Quantum Mechanics”, “Statistical Physics”, “Atomic Property”, “Atomic and Molecular Collision”, “Optics” and “Condensed Material”. There are 19,995, 11,080, 16,208, 11,941, 21,780, and 10,872 authors in those subdomains, respectively, with 106,611, 41,480, 116,888, 108,807, 154,392, and 52,134 papers. In table 2, we give detailed information regarding the APS dataset used in our experiments.

The PubMed dataset [54] consists of research articles containing metadata such as title, authors, source journal, predicted subfields, and publication dates. It can be accessed online by following the defined steps mentioned in PubMed url [2]. The dataset comprises 9,999 articles. It provides insights into scientific publications, including author collaborations and research trends across various biomedical domains. The dataset is structured to support bibliometric analysis, author network studies, and trend identification in academic publishing. The filtered dataset excludes papers without authors and standardizes the publication year format for consistency. We have identified three broad subdomains under Biology e.g. “Biochemistry”, “Cell Biology”, and “Genetics”. There are 2,517, 5,960, and 1,411 authors in those subdomains, respectively, with 2,531, 6,049, and 1,419 papers. In table 3, we give detailed information regarding the PubMed dataset used in our experiments.

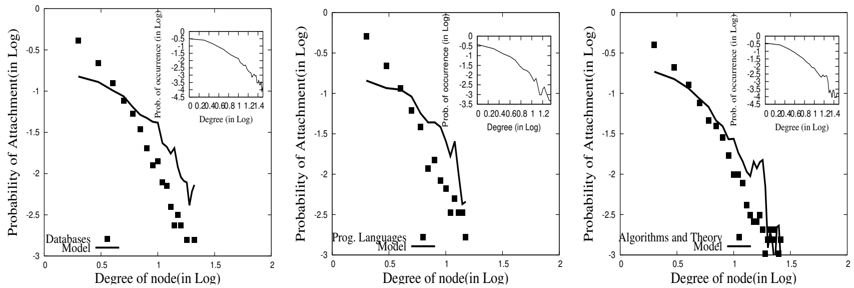
3.4 Preliminary Model of attachment

In this section, we empirically observe the attachment probability with respect to the social capital components and social ties of the nodes. The goal is to investigate whether, for each of these parameters, the known preferential attachment principle is followed during attachment.

[2] https://pubmed.ncbi.nlm.nih.gov/

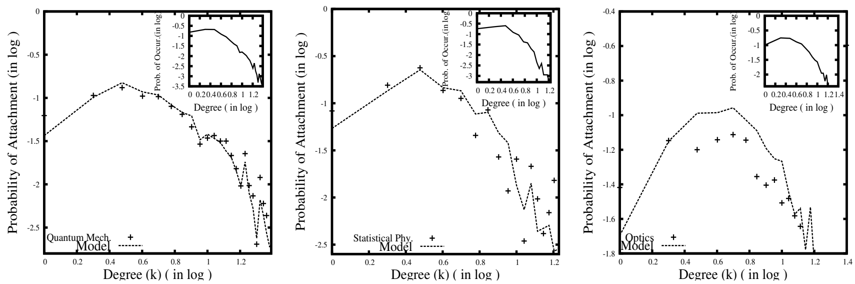
3.4.1 Attachment probability concerning structural capital

We attempt to analyze the importance of the structural capital of a node on the growth of collaboration networks. We observe the probability of attachment of a node to another node i with respect to its degree (d_i) and attempt to analytically model the same.



(a) Databases (b) Programming Languages (c) Algorithm & Theory

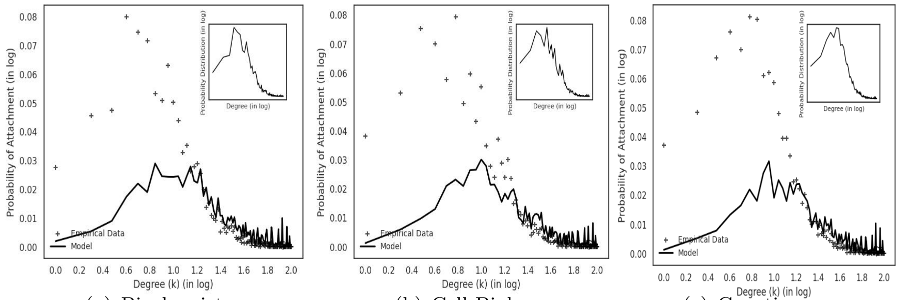
Fig. 3: It shows attachment Probability with respect to node degree D for "Databases", "Programming Languages" and "Algorithms & Theory" in figure 3(a), 3(b) and 3(c) respectively. In the inset, we show the degree distribution of the same network.



(a) Quantum Mech. (b) Statistical Physics (c) Optics

Fig. 4: It shows attachment Probability with respect to node degree D for "Quantum Mechanics", "Statistical Physics" and "Optics" in figure 4(a), 4(b) and 4(c) respectively. In the inset, we show the degree distribution of the same network.

Figure 3 shows the probability of attachment of a node to another node with respect to its degree for the three different subareas of CSE, whereas figure 4 shows the same for different sub-domains in Physics. We further observe the probability of attachment of a node to another node with respect to its degree



(a) Biochemistry (b) Cell Biology (c) Genetics

Fig. 5: It shows attachment Probability with respect to node degree D for "Biochemistry", "Cell Biology" and "Genetics" in figure 5(a), 5(b) and 5(c) respectively. In the inset, we show the degree distribution of the same network.

for the three different subareas of Biology in figure 5. We model the probability of a new attachment with respect to the node degree k (representing the structural capital of the node) as

$$A_k^{(S)} = \frac{k p_k^{(S)}}{\sum_j j p_j^{(S)}} \tag{3}$$

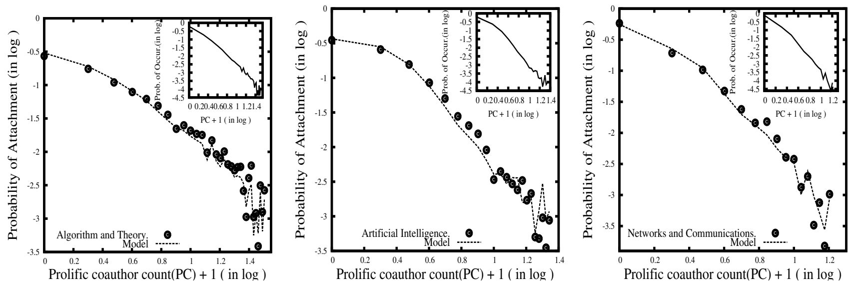
where $p_k^{(S)}$ represents the fraction of nodes with degree k. In figure 3 and 4, we also compare the empirically observed and the model values (given in equation 3) for the attachment probability of a link from a node to another node with respect to the degree D of the node. Figures 3(a), 3(b) and 3(c) compare the model values with the empirically observed values and corresponding RMSE values are 0.04, 0.06 and 0.04 for "Databases", "Programming Languages" and "Algorithms & Theory" respectively. Similarly, figures 4(a), 4(b) and 4(c) compare the model values with the empirically observed values and corresponding RMSE values are 0.008, 0.016 and 0.05 for "Quantum Mechanics", "Statistical Physics" and "Optics" respectively. Similarly, figures 5(a), 5(b) and 5(c) compare the model values with the empirically observed values and corresponding RMSE values are 0.013, 0.0125 and 0.014 for "Biochemistry", "Cell Biology" and "Genetics" respectively. Similar to the observations made through several previous studies [19, 24], the log-log plot represents an almost straight line for lower node degrees indicating a preferential attachment of the nodes. The RMSE values when compared with the empirical data ranges between 0.008 and 0.06 for the different sub-domains. Thus these findings indicate that structural capital plays a major role in new collaborations as nodes with high structural capital gain more collaborations as compared to others.

3.4.2 Attachment probability concerning relational capital

We investigate the importance of the relational capital of a node on the growth of collaboration networks. We specifically observe the attachment probability with respect to the prolific coauthor count of the nodes. We generate separate networks for the three sub-domains of CSE and Physics after identifying the prolific authors over a 4 year (2000 – 2003) time window. We labeled the authors as prolific when they published at least three papers in that time duration. We count the prolific coauthors in the neighborhood of each node and measure the attachment probability to the nodes with respect to their prolific coauthor count. Assuming a preferential attachment of the nodes, we further model the attachment probability with respect to the prolific author count l of the nodes (representing relational capital) as

$$A_l^{(R)} = \frac{l p_l^{(R)}}{\sum_j j p_j^{(R)}} \tag{4}$$

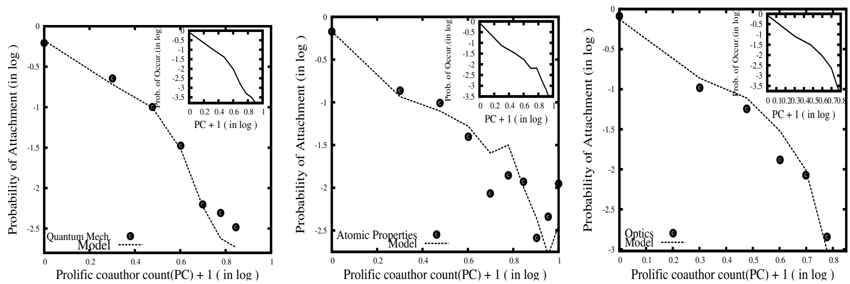
where $p_l^{(R)}$ represents the fraction of nodes with prolific coauthor count of l.



(a) Algorithm and Theory (b) Artificial Intelligence (c) Networks and Communications

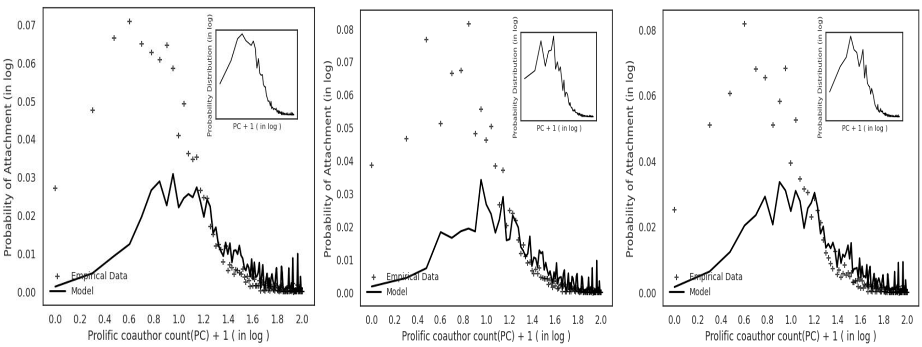
Fig. 6: It shows attachment Probability with respect to prolific coauthor count PC for "Algorithm and Theory", "Artificial Intelligence" and "Networks and Communications" in figures 6(a), 6(b) and 6(c) respectively. In the inset, we show the prolific coauthor count distribution of the same network.

Observations comparing the attachment probabilities derived using the model ($A_l^{(R)}$) with the empirically observed values with respect to l is shown in figures 6 and 7 for the different sub-domains of CSE and Physics. Figures 6(a), 6(b) and 6(c) compare the model values with the empirically observed values and corresponding RMSE values are 0.007, 0.007 and 0.01 for "Algorithm and Theory", "Artificial Intelligence" and "Networks and Communications" respectively. Similarly, figures 7(a), 7(b) and 7(c) compare the model values with the empirically observed values and corresponding RMSE values are 0.02, 0.012 and 0.033 for "Quantum Mechanics", "Atomic Properties" and "Optics"



(a) Quantum Mechanics (b) Atomic Properties (c) Optics

Fig. 7: It shows attachment Probability with respect to prolific coauthor count *PC* for "Quantum Mechanics", "Atomic Properties" and "Optics" in figures 7(a), 7(b) and 7(c) respectively. In the inset, we show the prolific coauthor count distribution of the same network.



(a) Biochemistry (b) Cell Biology (c) Genetics

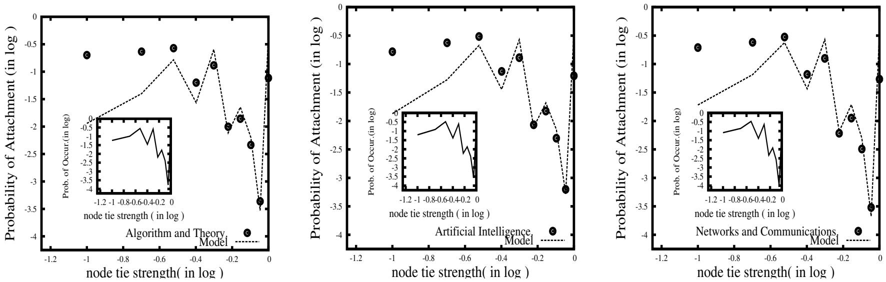
Fig. 8: It shows attachment Probability with respect to prolific coauthor count *PC* for "Biochemistry", "Cell Biology" and "Genetics" in figures 8(a), 8(b) and 8(c) respectively. In the inset, we show the prolific coauthor count distribution of the same network.

respectively. In figures 8(a), 8(b) and 8(c) that compare the model values with the empirically observed values and corresponding RMSE values are 0.013, 0.012 and 0.015 for "Biochemistry", "Cell Biology" and "Genetics" respectively. We observe the similar behavior as seen in various sub-domains of CSE and Physics. Significantly low RMSE values ranging from 0.007 to 0.03 for both the subject areas establish a good fit using the model. This further implies that the relational capital of the authors plays a role in new collaborations and authors with high relational capital are likely to have more collaborations as compared to the ones with lower relational capital.

We next observe the effect of social ties between coauthors in further collaborations.

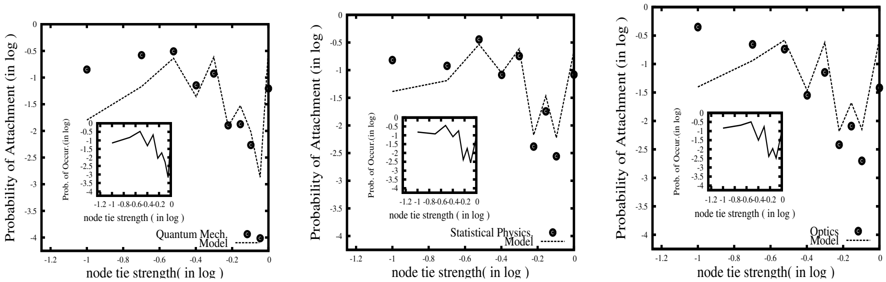
3.4.3 Attachment probability concerning social ties

In this section, we describe the role of social ties among the coauthors in further collaborations and publications. To observe the role of social ties in network evolution, we study whether the node tie strength also follows preferential attachment policy in republishing dynamics. We empirically validate this intuition and attempt to model the attachment probability with respect to the node tie strength.



(a) Algorithm and Theory. (b) Artificial Intelligence. (c) Networks and Communications.

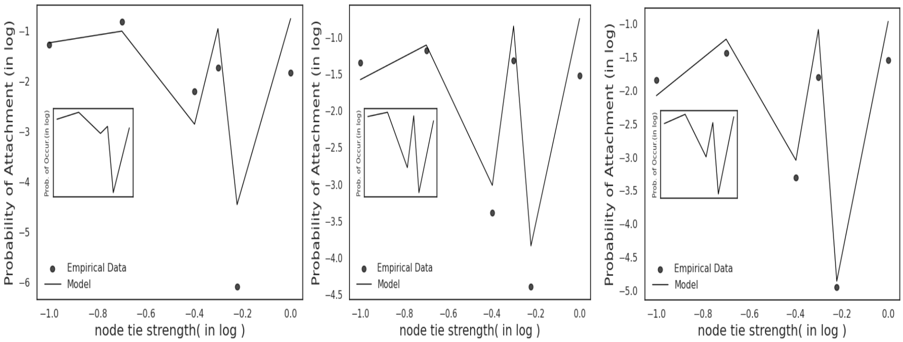
Fig. 9: It shows attachment Probability a node with respect to its node tie strength for "Algorithm and Theory", "Artificial Intelligence" and "Networks and Communications" in figure 9(a), 9(b) and 9(c)respectively. The inset shows the distribution of the node tie strength τ_i^m of the same network.



(a) Quantum Mechanics. (b) Statistical Physics. (c) Optics.

Fig. 10: It shows attachment Probability a node with respect to its node tie strength for "Quantum Mechanics", "Statistical Physics" and "Optics" in figures 10(a), 10(b) and 10(c) respectively. The inset shows the distribution of the node tie strength τ_i^m of the same network.

To investigate the influence of node tie strength on network growth, we focus on the reattachment of the already connected node pairs with respect to



(a) Biochemistry (b) Cell Biology (c) Genetics

Fig. 11: It shows attachment Probability a node with respect to its node tie strength for "Biochemistry", "Cell Biology" and "Genetics" in figures 11(a), 11(b) and 11(c) respectively. The inset shows the distribution of the node tie strength τ_i^m of the same network.

their node tie strength. Thus for every new publication between two existing coauthors, we observe their node tie strength and finally derive the *reattachment probability* with respect to these node tie strengths. To validate the existence of preferential attachment in such connections we model the reattachment probability, $A_t^{(T)}$, of already connected nodes with respect to the node tie strength t as

$$A_t^{(T)} = \frac{tp_t^{(T)}}{\sum_j jp_j^{(T)}} \tag{5}$$

where $p_t^{(T)}$ represents the probability of randomly selecting a node with tie strength t. Comparisons of $A_t^{(T)}$ (given in equation 5) and the empirically observed values for different sub-domains of CSE and Physics are shown in figures 9 and 10. Figures 9(a), 9(b) and 9(c) compare the model values with the empirically observed values and corresponding RMSE values are 0.011, 0.017 and 0.015 for "Algorithm and Theory", "Artificial Intelligence" and "Networks and Communications" respectively. Similarly, figures 10(a), 10(b) and 10(c) compare the model values with the empirically observed values and corresponding RMSE values are 0.03, 0.035 and 0.028 for "Quantum Mechanics", "Statistical Physics" and "Optics" respectively. Similarly, figures 11(a), 11(b) and 11(c) compare the model values with the empirically observed values and corresponding RMSE values are 0.099, 0.1 and 0.096 for "Biochemistry", "Cell Biology" and "Genetics" respectively. As can be observed, the model fits significantly well with the empirical values for moderate and high values of node tie strength.

Summary of the inferences: Thus in this section, we have modeled the attachment probability in the co-authorship networks with respect to the structural and relational capital as well as the social tie strength of the nodes. We discover that new attachments follow the preferential principle with respect to the structural and relational capital of the nodes whereas, reattachment due to edge formation between already connected nodes is preferential with respect to the node tie strengths. Thus these parameters determine the evolution and structure of co-authorship networks across different domains and hence need to be closely investigated.

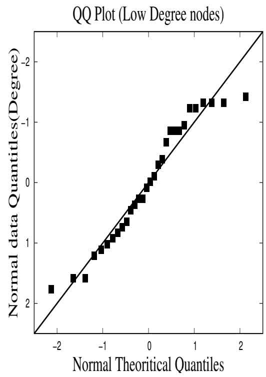
Hence we next investigate the role played by these social capital components along with the tie strength in the dynamics of link formation and evolution of the co-authorship network. Although we have performed all the further experiments in our work in three broad domains of CSE, Physics and Biology, we have reported only in the domains of CSE and Physics to reduce the redundant results. This reduction does not loose the generality of the work, on the other hand, it enhances the readability of the broader audience without overwhelming with such multiple results.

4 Experiment

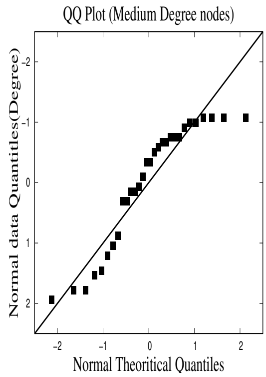
In this section, we investigate the role played by the social capital components and tie strength in the dynamics of link formation in co-authorship networks. We consider two major dynamics: a) emergence of new co-authorship links and b) publications between existing co-authors. While the first dynamics lead to the formation of new edges in the co-authorship network, the latter dynamics strengthen the edge weight between already connected nodes. Thus, both these dynamics contribute differently to the evolution of the co-authorship networks. Using a three-factor analysis of variance, we investigate whether there are significant main as well as interaction effects of the social capital components and node tie strength of the nodes on the mean number of new edges formed between both disconnected as well as already connected nodes. We initially outline the design of the experiments followed by an interpretation of the results obtained.

4.1 Design of ANOVA test

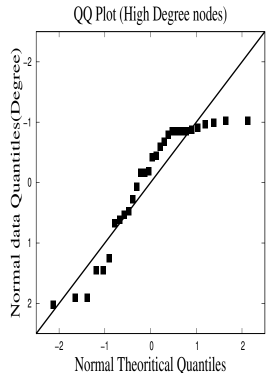
In this section, we outline the steps of the design of experiments. Initially, the co-authorship network is formed over a four-year period (2004 to 2007).To observe two dynamics separately, we consider publication count as the dependent variable and degree as structural capital (f_s), prolific co-author count as relational capital (f_r), and node tie strength (f_t) as independent variables of the ANOVA test. We observe the number of publications with existing co-authors over the next **two** years (2008 to 2009) at three different levels of each of the factors, the structural capital (f_s), relational capital (f_r) and node tie strength (f_t). For each of these factors, we categorize the levels as high (l_h), moderate (l_m) and low (l_o). Table 4 & 5 highlights the actual threshold values of the measures used for each of these factors for the level categorization for CSE and Physics domains respectively. The threshold values are determined based on a Head/Tail breaks algorithm proposed in [55].



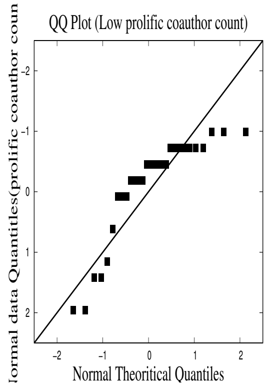
(a)



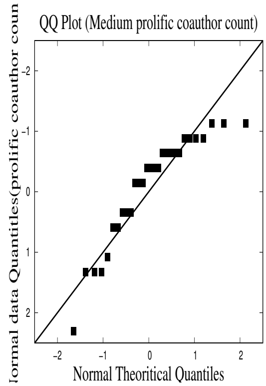
(b)



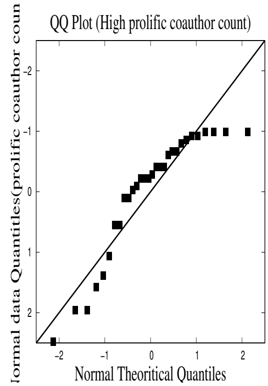
(c)



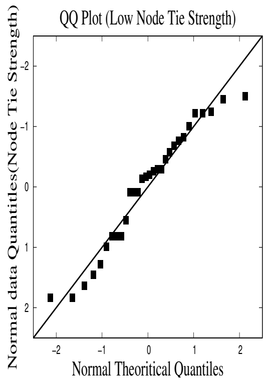
(d)



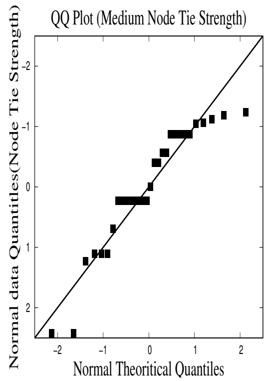
(e)



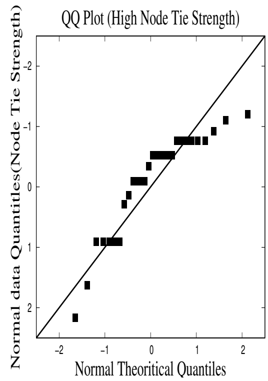
(f)



(g)



(h)



(i)

Fig. 12: The above figures show the quantile plot of degree, prolific coauthor count, and node tie strength in the domain of CSE. The x-axis represents Normal theoretical quantile values and y axis plots the sample quantile values of the corresponding parameter.

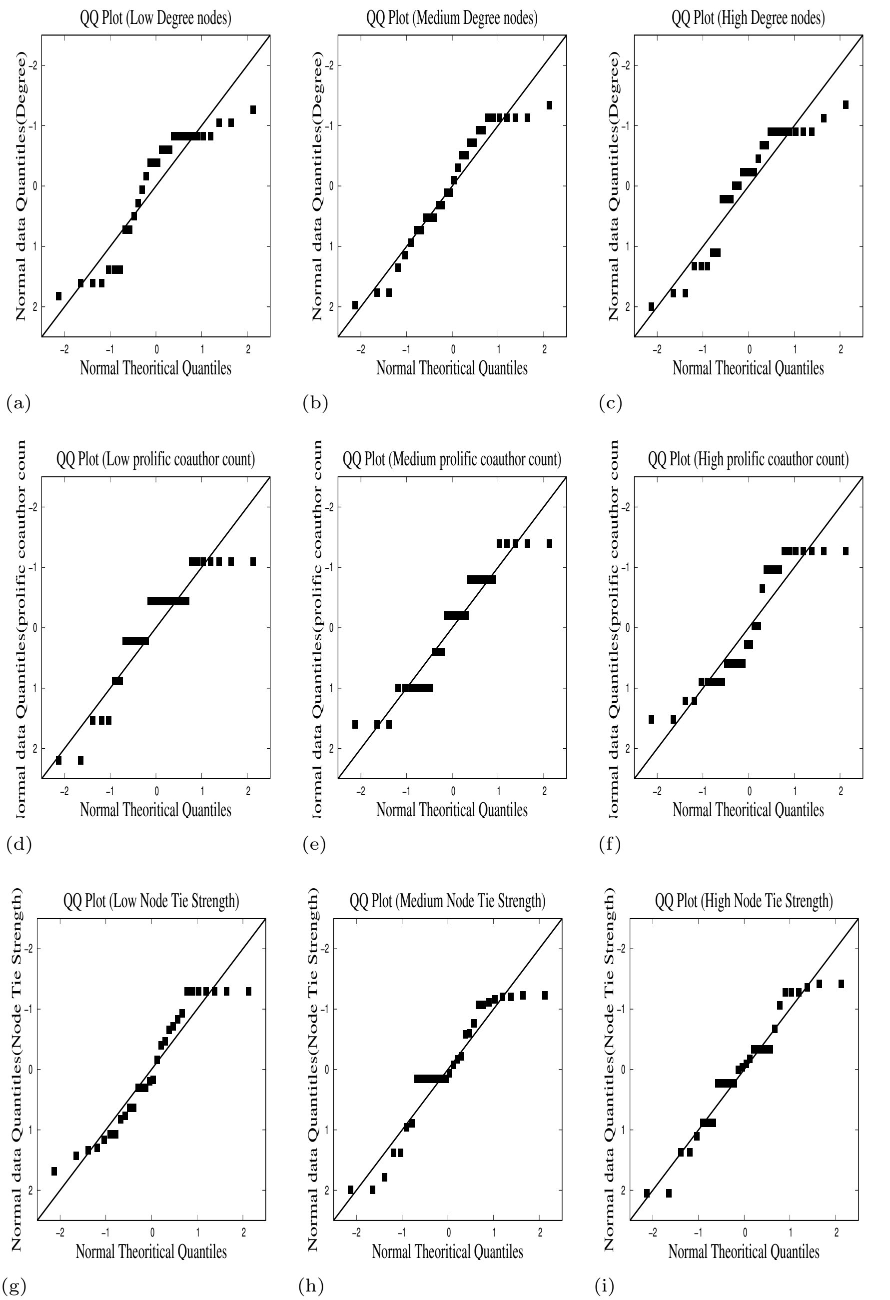


Fig. 13: The above figures show the quantile plot of degree, prolific coauthor count, and node tie strength in the domain of Physics. The x-axis represents Normal theoretical quantile values and the y-axis plots the sample quantile values of the corresponding parameter.

Table 4: Ranges of three levels of three factors in the field of CSE.

CSE	Low	Medium	High
Degree	1-40	40-80	80-225
Prolific coauthor count	0-16.5	16.5-33.75	33.75-88
Node tie strength	0-0.2	0.2-0.5	0.5-1

Table 5: Ranges of three levels of three factors in the field of Physics.

Physics	Low	Medium	High
Degree	1-16.5	16.5-32.55	32.55-96
Prolific coauthor count	0-5.5	5.5-11.5	11.5-23
Node tie strength	0-0.2	0.2-0.5	0.5-1

Table 6: Skewness of three factors in the field of CSE.

Skewness	Low	Medium	High
Degree	0.89	0.51	2.04
Prolific coauthor count	1.25	0.73	1.13
Node tie strength	0.304	0.796	1.022

We initially consider ($n = 30$) observations for each of the treatment combinations determined by the three levels of each factor f_s, f_r, and f_t. These observations were drawn randomly from the population of nodes belonging to each category. We initially do tests for basic preconditions of ANOVA i.e., a)normally distributed population and b)homogeneity of variances. To ensure the normality assumption of the data set, we use a *qq* plot to visually ensure the normality condition of the data shown in figure 12 & 13. We also derive the skewness and kurtosis coefficient (shown in table 6, 8, 7 & 9) to confirm the normality of the dataset.

The findings of skewness values of three social capital parameters in the CSE field reveal important insights as shown in table 6. A degree skewness of 2.04 indicates a right-skewed distribution, where a small number of researchers dominate collaborations, creating influential hubs that enhance knowledge sharing but may foster dependencies on key individuals. The moderate to high skewness in prolific co-author counts shows that while most researchers collaborate moderately, a few significantly outpace their peers, highlighting the critical role of prolific authors in driving research output. Additionally, the skewness in node tie strength suggests a mix of weak and strong ties, with many researchers maintaining weak connections and some fostering stronger relationships. This variation is vital, as strong ties facilitate deeper collaboration and knowledge transfer, influencing idea diffusion and the capacity for diverse collaborations. Overall, these findings underscore the structured nature of collaboration in CSE, characterized by influential hubs, prolific authors, and a balance of weak and strong ties.

The analysis of kurtosis(shown in table 7) reveals distinct patterns in the collaboration dynamics of researchers in CSE domain. The high kurtosis value

Table 7: Kurtosis of three factors in the field of CSE.

Kurtosis	Low	Medium	High
Degree	-0.44	-0.74	4.81
Prolific coauthor count	0.65	-0.49	0.28
Node tie strength	-0.821	0.353	0.206

Table 8: Skewness of three factors in the field of Physics.

Skewness	Low	Medium	High
Degree	-0.04	-0.20	2.28
Prolific coauthor count	0.87	0.14	0.50
Node tie strength	0.89	0.55	0.36

Table 9: Kurtosis of three factors in the field of Physics.

Kurtosis	Low	Medium	High
Degree	-1.185	-0.65	5.75
Prolific coauthor count	-0.14	-0.93	-0.48
Node tie strength	0.56	-0.55	-0.50

of 4.81 for the degree of collaboration indicates a heavy-tailed distribution, suggesting that while most researchers have few connections, a small number exhibit very high levels of collaboration, creating outlier individuals who dominate the network. In contrast, the kurtosis values for prolific co-author count are close to zero, indicating a distribution that resembles normality; this suggests that the majority of researchers collaborate moderately, with few extreme values, resulting in a balanced and evenly distributed collaboration pattern. Additionally, the low kurtosis value of -0.821 for node tie strength indicates lighter tails, implying that most researchers have relatively weak ties and fewer instances of strong connections. This points to a more uniform distribution of tie strengths within the collaboration network, lacking extreme cases of very strong ties. Overall, these findings highlight the diverse characteristics of collaboration dynamics, emphasizing the presence of influential outliers, a balanced co-authorship distribution, and predominantly weak ties among collaborators.

The result of skewness and kurtosis observed in the above-referred table indicates that in all cases, data is normally distributed as the values of skewness and Kurtosis are within the standard range of reference for normality condition. Thus, sample data for three factors pass the precondition of normality for the ANOVA test.

We subsequently test the homogeneity of variances across each of the categories for all the factors in the domain of CSE and Physics at 0.05 level of significance using Levene's test for homogeneity of variances (results shown in the table 10 & 11). The Head/Tail break divides data in such a way that the high category of each factor has a long range. To keep the data ranges normalized with other categories of high category of each factor, we kept our sample in the following ranges: High degree (80 – 120-(CSE), 32.5 – 49-(Physics)), High prolific coauthor count (33.75 – 50.00-(CSE), 11.5 – 17-(Physics)), High node tie strength (0.5 – 0.8-(CSE), 0.5 – 0.8-(Physics)). In the result of Levene's test, we estimate the mean and standard deviation (SD) of each factor in every three categories. The result shows that though there is a significant difference among the mean values of each category, the standard deviation of each category differs very little. Thus, the results infer that there is no difference in variance ($p > 0.05$) of the data across all the factors in both the CSE and Physics domains, thus it maintains the precondition of homogeneity of variances for ANOVA test.

Table 10: Levene's test result in the domain of CSE.

CSE	Category	Mean	Std. Deviation(SD)	F	p
Degree	Low	16.133	10.66	0.117	0.889
	Medium	54.167	12.259		
	High	95.167	11.568		
Prolific Coauthor count	Low	3.7	3.734	1.825	0.167
	Medium	21.6	4.073		
	High	40.267	4.73		
Node Tie Strength	Low	0.133	0.055	1.820	0.165
	Medium	0.367	0.076		
	High	0.53	0.075		

Table 11: Levene's test result in the domain of Physics.

Physics	Category	Mean	Std. Deviation (SD)	F	p
Degree	Low	5.93	3.44	0.42	0.66
	Medium	24.6	4.05		
	High	41.27	4.14		
Prolific Coauthor count	Low	1.46	1.41	0.938	0.39
	Medium	8.27	1.76		
	High	16.16	1.60		
Node Tie Strength	Low	0.099	0.056	0.52	0.60
	Medium	0.34	0.086		
	High	0.646	0.067		

We next outline the ANOVA results and the interpretation of the results.

5 Results and Discussions

In this section, we first observe the result of ANOVA for two distinct dynamics of collaboration formation, i.e., a) formation of the new edge between two nodes and b) increment of the weight of the existing edge. We observe the effect of three factors i.e., degree, prolific coauthor count, and node tie strength in those dynamics in two different domains of research i.e., CSE and Physics separately. We further interpret the result of ANOVA for each different domain and dynamics to explain the main effect as well as interaction effects among the factors.

5.1 Effects on the new edge formation dynamics

In the domain of CSE, the main effect of the factors,i.e., degree, prolific coauthor count, and node tie strength are significant ($p < 0.05$). The significant p-value indicates that the mean responses for three different levels of categorization, i.e., 'High', 'Medium', and 'Low' of each of the factors are significantly different. This nature of the main effect reveals that different levels of these three factors have a distinct amount of impact on publication count. To observe the interaction among those three factors and to understand their interplay in new collaboration development, we perform the pair-wise interaction among the factors. The p-value for the interaction between node tie strength and prolific coauthor count and interaction between degree and node tie strength are 0.2028 and 0.1936, are not significant which indicates that the effect of the degree and prolific coauthor count (on the number of publication) having an insignificant dependency on node tie strength. Whereas interaction between the degree and prolific coauthor count is 0 which indicates the effect of the degree of the node depends on the prolific coauthor count.

To understand the interaction between degree and prolific coauthor count in new edge formation precisely, we draw the interaction diagram. In figure 14(a), we observe the interaction between degree and prolific coauthor count. We consider random 30 data points of each category. The variation of the slope of the three lines indicates the significant interaction between degree and prolific coauthor count. We observe that the average number of publication counts not only depends on degree but also depends on prolific coauthor count in case of new edge formation. It is prominently observable that the number of publication counts of high degree nodes depends on the prolific coauthor count. Thus, it reveals that authors who have many collaborations and out of those collaborators have a large set of prolific coauthors who produce more publications. Compared to that, authors having a higher number of collaborators but few of them having a higher number of prolific coauthors produce less number of publications. On the other hand, authors having a medium and low degree, but having a medium number of prolific coauthors comparatively publish a higher number of papers than that of the authors having a higher and lower number of prolific coauthors. This behavior could be explained by authors already having a higher number of prolific coauthors possibly getting less opportunity to get involved in new collaboration over a publication, on the other hand, authors having medium prolific coauthor count undergo more publication through a new collaboration. In summary, we get that in the domain of CSE, the development of new collaboration eventually depends on the degree and prolific coauthor count, and hardly a significant effect of node tie strength is observable. This nature of new collaboration development in the

CSE domain, explains the dominance of structural and relational capital over social tie strength.

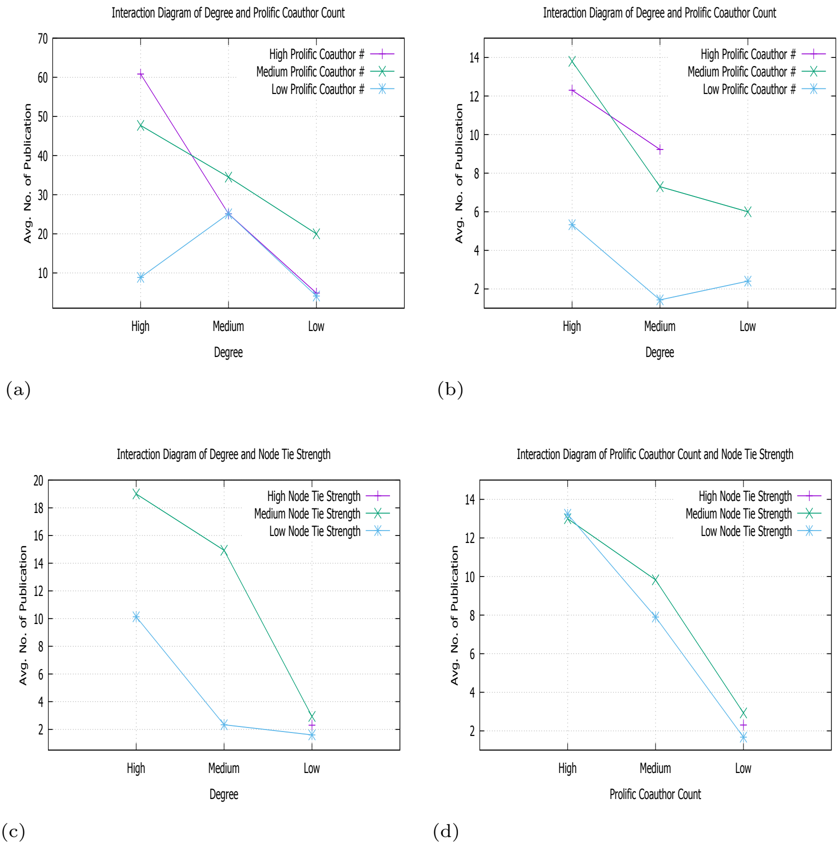


Fig. 14: Figure 14(a) shows the interaction between degree and prolific coauthor count in the domain of CSE in the case of new link formation dynamics. Similarly, in fig. 14(b), (c) and (d) show the pairwise interaction between degree and prolific coauthor count, degree and node tie strength, prolific coauthor count and node tie strength respectively in the domain of Physics in the case of new link formation dynamics.

We also observe the effect of three factors in new edge formation in the domain of Physics separately. The main effect of the factors,i.e., degree, prolific coauthor count and node tie strength are significant ($p < 0.05$). The significant p-value indicates that the mean effect of three different levels of categorization, i.e., 'High', 'Medium', and 'Low' of each of the factors are significantly different on the development of new collaboration formation in the domain of Physics. We further observe that in the domain of physics pairwise interaction

(p values) among the three factors are $0.0069, 0$, and 0.045 respectively which are significant. It reveals that there is a significant amount of dependency of other factors on each factor in contributing to new collaboration development.

To observe the interplay of three factors in new collaboration development more precisely, we draw the interaction diagram of each pair of factors separately. In figure14(b), we observe the interaction between degree and prolific coauthor count in the domain of physics. We took the average of publication count of 30 random data samples of each category, though we did not find any data points for the low degree, high prolific coauthor count which is quite obvious. The variation of the slope of the three lines indicates the interaction between them. In the Physics domain, we observe a similar interaction between degree and prolific coauthor count in the CSE domain and the nature of the interaction could be explained earlier. But interestingly, in the domain of physics, in new collaboration development dynamics, node tie strength has a significant role which can be observable in the figure14(c) & (d). In figure14(b), we observe that authors who have a higher number of collaborators and with a higher number of collaborators who have collaborated with prolific authors undergo a lesser number of new publications. On the other hand, authors produce more new publications with those authors who have a comparatively lesser number of collaborations with prolific authors. This situation could be explained as that the authors having many prolific coauthors can have less chance to undergo a new collaboration. On the other hand, in figure14(c)&(d), we observe that authors having medium node tie strength and high degree/ high prolific coauthor count have a comparatively higher number of publications than other categories. Though the node tie strength does not dominate over the degree and prolific coauthor count in a physics domain, node tie strength has a significant role in new collaboration development. These phenomena could be explained as follows– selection criteria of new collaborators not only on the degree and prolific coauthor count but also their social tie strength among their existing collaborators. This justifies the growth of the Physics domain differently from that of the CSE domain.

5.2 Effects on republishing dynamics

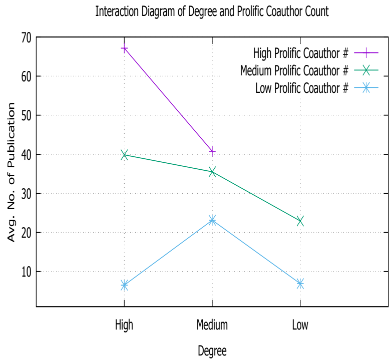
We observe the effect of three factors in republishing dynamics in the domain of CSE and Physics separately. In the domain of CSE, the main effect of the factors, i.e., degree, prolific coauthor count and node tie strength are significant ($p < 0.05$). The significant p-value indicates that the mean responses for three different levels of categorization, i.e., 'High', 'Medium', and 'Low' of each of the factors are significantly different. The p-value for the interaction between degree and prolific coauthor count and interaction between degree and node tie strength are 0 and 0.0063, are significant which indicates the significant evidence that the effect of the degree depends on node tie strength and prolific coauthor count. Whereas interaction between the node tie strength and prolific coauthor count is 0.6299 which indicates the very trivial effect of prolific coauthor count on node tie strength.

To observe the interaction between degree and prolific coauthor count in republishing dynamics, we plot the interaction diagram which is shown in fig15(a). We observe that for high-degree nodes that undergo a new publication with their existing collaborators in the domain of CSE, preferably select among the existing collaborators who have highly prolific coauthors in their neighborhood. It signifies that preference for the selection of the existing collaborator is highly influenced by the existence of prolific coauthors in their neighborhood. We also observe the interaction between degree and node tie strength in the domain of CSE in fig.15(b). We do not find those data points that correspond to high degree-high node tie strength and high degree-medium node tie strength which are quite obvious. Variation of the slopes of the lines indicates the interaction between degree and node tie strength. In fig.15(b), we observe that medium degree node with medium node tie strength value has a significantly high number of an average number of publications. This nature could be explained as republishing with existing collaborators affected by node tie strength and prolific coauthor count. In the domain of CSE, an author is preferably selected among the existing collaborators by their number of prolific coauthors and the value of node tie strength.

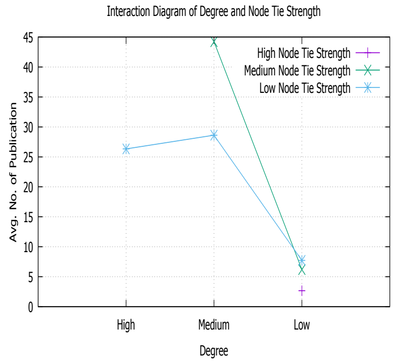
Similarly, we observe the effect of three factors in republishing dynamics in the domain of Physics separately. The main effect of the factors,i.e., degree, prolific coauthor count and node tie strength are significant ($p < 0.05$). The significant p-value indicates that the mean responses for three different levels of categorization, i.e., 'High', 'Medium', and 'Low' of each of the factors are significantly different. This nature of the main effect reveals that different levels of these three factors have a distinct amount of impact on publication count which shows similar phenomena as new collaboration dynamics. The p-value for the interaction between degree and prolific coauthor count and interaction between degree and node tie strength are 0.0001 and 0, are significant which indicates the significant evidence that the effect of the degree depends on node tie strength and prolific coauthor count. Whereas interaction between the node tie strength and prolific coauthor count is 0.5729 which indicates the insignificant interaction between prolific coauthor count and node tie strength.

In republishing dynamics, in the domain of CSE, prolific coauthors have a significant role. we observe that though the degree of the node is high, but being low prolific coauthor count avg. the number of publication count is significantly low. On the other hand, authors having a high prolific coauthor count produce a comparatively higher number of publications. This behavior possibly be explained as follows——For a new publication in CSE domain, the author prefers to collaborate with an existing collaborator who has a high number of prolific coauthors in his neighborhood.

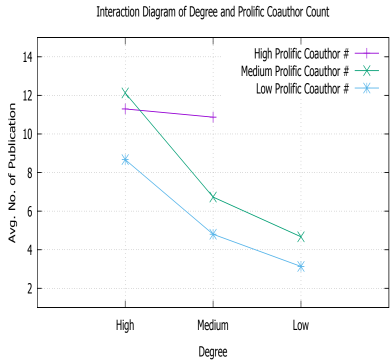
This nature of republishing dynamics differs in the Physics domain, we observe in figure15(c) that there is an interaction between degree and prolific coauthor count. For a new publication, the author prefers to collaborate with an author having a medium number of prolific coauthors. This nature of republishing dynamics can be explained as follows—- The authors having



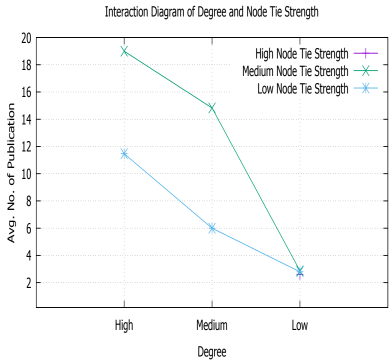
(a)



(b)



(c)



(d)

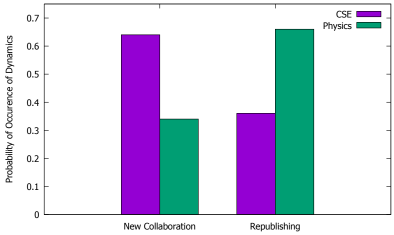
Fig. 15: Figures 15(a) and (b) show the interaction between degree and prolific coauthor count, degree and node tie strength respectively in the case of republishing dynamics in the domain of CSE. Similarly, in fig. 15(c) and (d) show the interaction between degree and prolific coauthor count, degree and node tie strength respectively in the case of republishing dynamics in the domain of Physics.

more number of prolific coauthors get less opportunity to collaborate over a new publication with their existing collaborators. Thus in this domain, authors having a medium prolific coauthor count get more chances to collaborate on a new publication among their existing collaborators. This nature reveals the dynamics of close group research work.

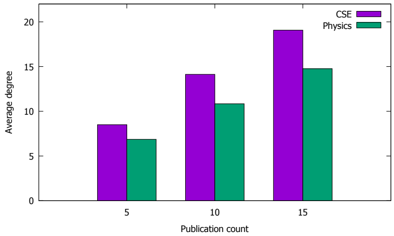
On the other hand, the interaction diagram (15(b) & (d)) of node tie strength and degree in the domain of CSE & Physics show a similar nature. From both of the diagrams, we observe that an average number of publication count is affected by node tie strength. It can be explained as—- authors preferably attach to the existing collaborators having medium node tie strength than that of the existing collaborators having low node tie strength. In both of the domains, there is a dominant role of node tie strength in republishing dynamics. Thus it can be observed that social tie strength with other collaborators influences the selection of the collaborator for a new publication.

Thus, we observe the significant difference between the republishing dynamics in these two domain which possibly explain the topological difference between the CSE and Physics domain.

5.3 Dominance of social capital and social ties



(a) Variation of Dynamics across the domains of CSE and Physics.



(b) Variation of avg.degree across the domains CSE and Physics.

Fig. 16: Figure 16(a) represents the variation of probability of occurrence of two distinct dynamics i.e., new collaboration and republishing in the domain of "CSE" and "Physics". Figure 16(b) represents the variation of average degree when the authors have same number of publications in the different domains of research.

Now, we observe the varying role of social capital and tie strength in network evolution to reveal the significant difference in the network topology across several research domains. In the earlier section, we observe the preferential attachment of nodes with respect to degree centrality and prolific coauthor count which reveals the nature of new link formation under the influence of structural and relational capital. In the last section, We observe that the two aspects of social capital contribute to new link formation dynamics, whereas tie strength among the coauthors contributes to republishing dynamics. To observe varying occurrences of these two evolution dynamics, we specifically observe the probability of occurrence of two evolution dynamics across two domains, e.g., CSE and Physics. To validate the outcome of the last section, We further observe varying dominance of social capital and social ties in the network evolution process across several domains.

Variation of evolution dynamics across domains In figure 16(a), it can be observed that the fraction of links formed due to two different evolution dynamics is significant and there exists a significant variation of the probability of occurrence of each dynamics across these two domains (CSE and Physics).

This variation of evolution dynamics across different domains could explain the topological difference among the various domains of research. The variation of the probability of occurrence of the different evolutionary dynamics in different domains further cross-validates the varying dominance of social capital and social ties in the network evolution process across several domains. We observe that in the Physics domain "republishing" dynamics is more probable than that of new collaboration formation. This nature of publishing reflects their interest in working in a close group. Thus, this could explain the causes of the lower value of Spearman rank correlation between degree centrality and publication count of Physics compared to CSE which has been observed in section 3.

Variation of number of collaborators To understand the variation of effects of social capital and social ties across the domains, we further measure the average number of collaborators of the authors in a domain those who have the same publication count. The result is highlighted in figure 16(b) which shows that the average degree for the same number of publication counts in CSE is comparatively higher than that of Physics. This nature of the collaboration pattern reveals the dominance of structural capital in CSE over social ties. On the other hand, in Physics, a low average degree infers the existence of a small closed group which might contribute to republishing dynamics among the existing collaborators. Thus, this observation also cross-validates the dominance of social tie strength in the domain of Physics.

This result validates the objective of the work i.e., the several social capital components as well as the node tie strength and their interplay could explain the existence of topological differences across the domain despite similar preferential attachment activities across the various domains and their sub-domains.

6 Conclusion

In this paper, we examine how social capital and the strength of social ties among coauthors affect the development of scientific collaboration networks. In the CSE domain, new collaborations are largely driven by an individual's network degree and prolific co-authors, with limited influence from social tie strength, highlighting the dominance of structural and relational capital. Conversely, in physics, node tie strength plays a significant role in forming new collaborations, showing that the importance of social ties varies across disciplines. Interestingly, both fields show a dominant role for node tie strength in republishing dynamics. Physics exhibits a higher likelihood of republishing among small, closed groups of collaborators, whereas CSE's higher average degree reflects a broader, more open collaboration network. This comparison underscores the differing patterns of collaboration between the two domains, with CSE favoring structural capital and physics emphasizing strong social ties in certain aspects. This variation helps explain the topological differences observed among various research fields. Our findings may aid in the development of a collaboration recommendation system within the scientific research community and improve the assessment of researchers across different domains. The limitations of this work could be as follows—- a) lack of an explicit measure of cognitive capital and its detailed analysis in the context of collaboration formation b) we could not explore the influence of other factors e.g., institutional policy, individual motivation, etc. in collaboration establishment and further activities due to the lack of availability of data c) though the strong existence of 'preferential attachment', there might be other mechanisms in collaboration formation. We plan to work in that direction in the future which might explore the detailed understanding of cognitive capital in collaboration formation in scientific domains, the role of institutional policy in the development of new collaborations as well as activities on existing collaborations. Although we have modeled collaboration development influenced by social capital and tie strength, there is a need for a comprehensive model that captures collaboration network growth by considering the varying effects of these factors across different research domains, which we aim to develop in the future.

Declarations

- **Funding**: No funds, grants, or other support were received to conduct this study.
- **Conflict of interest/Competing interests**: On behalf of all authors, the corresponding author states that there is no conflict of interest.
- **Data availability**: Publicly available.
- **Author contribution**: **Avijit Gayen:** Conceptualization, Methodology, Experiment, Writing. **Somyajit Chakraborty:** Data collection, Experiment. **Nilotpal Chakraborty:** Writing review & editing. **Angshuman Jana:** Supervision, Writing review & editing.

References

[1] S.S. Moghadam, M.R. Gholamian, R. Zahedi, M. Shaqaqifar, Designing a multi-purpose network of sustainable and closed-loop renewable energy supply chain, considering reliability and circular economy. Applied Energy **369**, 123,539 (2024)

[2] R. Zahedi, S. Gitifar, A. Ahmadi, Optimum design of wind turbine, photovoltaic panel, diesel generator hybrid system and its integration with the distribution network. Journal of Renewable and New Energy (2024)

[3] A. Abbasi, R.T. Wigand, L. Hossain, Measuring social capital through network analysis and its influence on individual performance. Library & Information Science Research **36**(1), 66–73 (2014)

[4] Q. Gui, C. Liu, D. Du, Does network position foster knowledge production? evidence from international scientific collaboration network. Growth and Change **49**(4), 594–611 (2018)

[5] I. Kawachi, L.F. Berkman, Social ties and mental health. Journal of Urban Health **78**(3), 458–467 (2001)

[6] A.W. Pearson, J.C. Carr, J.C. Shaw, Toward a theory of familiness: A social capital perspective. Entrepreneurship theory and practice **32**(6), 949–969 (2008)

[7] W.W. Powell, D.R. White, K.W. Koput, J. Owen-Smith, Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences1. American journal of sociology **110**(4), 1132–1205 (2005)

[8] G. Ahuja, Collaboration networks, structural holes, and innovation: A longitudinal study. Administrative science quarterly **45**(3), 425–455 (2000)

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

[10] L. Zhai, X. Yan, J. Shibchurn, X. Song, Evolutionary analysis of international collaboration network of chinese scholars in management research. Scientometrics **98**(2), 1435–1454 (2014)

[11] A. Agrawal, A. Goldfarb, How do communication costs affect scientific collaboration? exploring the effect of bitnet. Unpublished draft, University of Toronto (2005)

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

[13] W. Tsai, S. Ghoshal, Social capital and value creation: The role of intrafirm networks. Academy of management Journal **41**(4), 464–476 (1998)

[14] M. Coccia, L. Wang, Evolution and convergence of the patterns of international scientific collaboration. Proceedings of the National Academy of Sciences **113**(8), 2057–2061 (2016)

[15] S.H. Chang, The evolutionary growth estimation model of international cooperative patent networks. Scientometrics pp. 1–19 (2017)

[16] M.E. Newman, Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences **101**(suppl 1),

5200–5205 (2004)

[17] L.C. Ribeiro, M.S. Rapini, L.A. Silva, E.M. Albuquerque, Growth patterns of the network of international collaboration in science. Scientometrics **114**(1), 159–179 (2018)

[18] C.K. Singh, S. Jolad, Structure and evolution of indian physics co-authorship networks. Scientometrics **118**(2), 385–406 (2019)

[19] H. Jeong, Z. Néda, A.L. Barabási, Measuring preferential attachment in evolving networks. EPL (Europhysics Letters) **61**(4), 567 (2003)

[20] S. Redner, Citation statistics from more than a century of physical review. Phys. Today **58**, 49–54 (2004)

[21] D. Lee, K.I. Goh, B. Kahng, D. Kim, Complete trails of coauthorship network evolution. Physical Review E **82**(2), 026,112 (2010)

[22] A.L. Barabási, R. Albert, Emergence of scaling in random networks. science **286**(5439), 509–512 (1999)

[23] A.L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations. Physica A: Statistical mechanics and its applications **311**(3), 590–614 (2002)

[24] R. Albert, A.L. Barabási, Topology of evolving networks: local events and universality. Physical review letters **85**(24), 5234 (2000)

[25] A. Barrat, M. Barthélemy, A. Vespignani, Weighted evolving networks: coupling topology and weight dynamics. Physical review letters **92**(22), 228,701 (2004)

[26] A. Barrat, M. Barthélemy, A. Vespignani, Modeling the evolution of weighted networks. Physical review E **70**(6), 066,149 (2004)

[27] M. Li, J. Wu, D. Wang, T. Zhou, Z. Di, Y. Fan, Evolving model of weighted networks inspired by scientific collaboration networks. Physica A: Statistical Mechanics and its Applications **375**(1), 355–364 (2007)

[28] X.L. Xu, C.P. Liu, D.R. He, A collaboration network model with multiple evolving factors. Chinese Physics Letters **33**(4), 048,901 (2016)

[29] M. Inoue, T. Pham, H. Shimodaira, Joint estimation of non-parametric transitivity and preferential attachment functions in scientific co-authorship networks. Journal of Informetrics **14**(3), 101,042 (2020)

[30] P. Leifeld, Polarization in the social sciences: Assortative mixing in social science collaboration networks is resilient to interventions. Physica A:

Statistical Mechanics and its Applications **507**, 510–523 (2018)

[31] Y. Zhao, R. Zhao, An evolutionary analysis of collaboration networks in scientometrics. Scientometrics **107**(2), 759–772 (2016)

[32] F. Battiston, J. Iacovacci, V. Nicosia, G. Bianconi, V. Latora, Emergence of multiplex communities in collaboration networks. PloS one **11**(1), e0147,451 (2016)

[33] M.V. Tomasello, G. Vaccario, F. Schweitzer, Data-driven modeling of collaboration networks: a cross-domain analysis. EPJ Data Science (2017). URL https://doi.org/10.1140/epjds/s13688-017-0117-5

[34] K. Sherif, J. Hoffman, B. Thomas, Can technology build organizational social capital? the case of a global it consulting firm. Information & management **43**(7), 795–804 (2006)

[35] W.S. Chow, L.S. Chan, Social network, social trust and shared goals in organizational knowledge sharing. Information & management **45**(7), 458–465 (2008)

[36] J. Lu, J. Yang, C.S. Yu, Is social capital effective for online learning? Information & Management **50**(7), 507–522 (2013)

[37] B. Ortiz, M.J. Donate, F. Guadamillas, Relational and cognitive social capital: Their influence on strategies of external knowledge acquisition. Procedia Computer Science **99**, 91–100 (2016)

[38] R.D. Putnam, et al., The strange disappearance of civic america. Policy: A Journal of Public Policy and Ideas **12**(1), 3 (1996)

[39] J. Nahapiet, S. Ghoshal, in *Knowledge and social capital* (Elsevier, 2000), pp. 119–157

[40] S. Madhavaram, R. Appan, K. Manis, G.J. Browne, Building capabilities for software development firm competitiveness: The role of intellectual capital and intra-firm relational capital. Information & Management **60**(2), 103,744 (2023)

[41] R.S. Burt, in *Social capital* (Routledge, 2017), pp. 31–56

[42] M. Steinmo, E. Rasmussen, The interplay of cognitive and relational social capital dimensions in university-industry collaboration: Overcoming the experience barrier. Research Policy **47**(10), 1964–1974 (2018)

[43] J. Hahn, J.Y. Moon, C. Zhang, Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties. Information Systems Research **19**(3), 369–391 (2008)

[44] Y.A. de Montjoye, A. Stopczynski, E. Shmueli, A. Pentland, S. Lehmann, The strength of the strongest ties in collaborative problem solving. Scientific reports **4**, 5277 (2014)

[45] S.L. Jack, The role, use and activation of strong and weak network ties: A qualitative analysis. Journal of management studies **42**(6), 1233–1259 (2005)

[46] R. Cross, J.N. Cummings, Tie and network correlates of individual performance in knowledge-intensive work. Academy of management journal **47**(6), 928–937 (2004)

[47] M.E. Sosa, Where do creative interactions come from? the role of tie content and social networks. Organization Science **22**(1), 1–21 (2011)

[48] K. Rost, The strength of strong ties in the creation of innovation. Research policy **40**(4), 588–604 (2011)

[49] Y. Lu, D. Yang, Information exchange in virtual communities under extreme disaster conditions. Decision Support Systems **50**(2), 529–538 (2011)

[50] T. Fushimi, K. Saito, T. Ikeda, K. Kazama, Estimating node connectedness in spatial network under stochastic link disconnection based on efficient sampling. Applied Network Science **4**(1), 1–24 (2019)

[51] R.A. Muniady, A.A. Mamun, M.R. Mohamad, P.Y. Permarupan, N.R.B. Zainol, The effect of cognitive and relational social capital on structural social capital and micro-enterprise performance. SAGE Open **5**(4), 2158244015611,187 (2015)

[52] T. Chakraborty, S. Sikdar, V. Tammana, N. Ganguly, A. Mukherjee, in *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on* (IEEE, 2013), pp. 426–433

[53] Aps, data sets for research. (2012) (Accessed 16th June, 2016)

[54] J. Kans, in *Entrez programming utilities help [Internet]* (National Center for Biotechnology Information (US), 2024)

[55] B. Jiang, Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. The Professional Geographer **65**(3), 482–494 (2013)