REGULAR ARTICLE

Open Access

Sustainability of Stack Exchange Q&A communities: the role of trust

Ana Vranić[1*] , Aleksandar Tomašević[2], Aleksandra Alorić[1,3] and Marija Mitrović Dankulov[1]

*Correspondence: anav@ipb.ac.rs
[1] Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, Belgrade, Serbia
Full list of author information is available at the end of the article

Abstract

Knowledge-sharing communities are fundamental elements of a knowledge-based society. Understanding how different factors influence their sustainability is of crucial importance. We explore the role of the social network structure and social trust in their sustainability. We analyze the early evolution of social networks in four pairs of active and closed Stack Exchange communities on topics of physics, astronomy, economics, and literature and use a dynamical reputation model to quantify the evolution of social trust in them. In addition, we study the evolution of two active communities on mathematics topics and two closed communities about startups and compare them with our main results. Active communities have higher local cohesiveness and develop stable, better-connected, trustworthy cores. The early emergence of a stable and trustworthy core may be crucial for sustainable knowledge-sharing communities.

Keywords: Networks structure; Dynamic reputation; Knowledge exchange; Stack Exchange; Sustainability of Q&A communities

1 Introduction

The development of a knowledge-based society is one of the critical processes in the modern world [1, 2]. In a knowledge-based society, knowledge is generated, shared, and made available to all members. It is a vital resource. Sharing this resource between individuals and organizations is a necessary process, and knowledge-sharing communities are one of the fundamental elements of a knowledge society.

Often, these knowledge-sharing communities depend on the willingness of their members to engage in an exchange of information and knowledge. Participation in the community is voluntary, with no noticeable material gains for members. Recent research has shown that the process of knowledge and information exchange is strongly influenced by *trust* [3, 4]. The exchange of knowledge depends on trust between a member and the community. It is a collective phenomenon that depends on and is built through social interactions between community members. This is why we believe it is crucial to understand how trustworthy knowledge-sharing communities emerge and disappear, as well as to unveil the fundamental mechanisms that underlie their evolution and determine their sustainability.
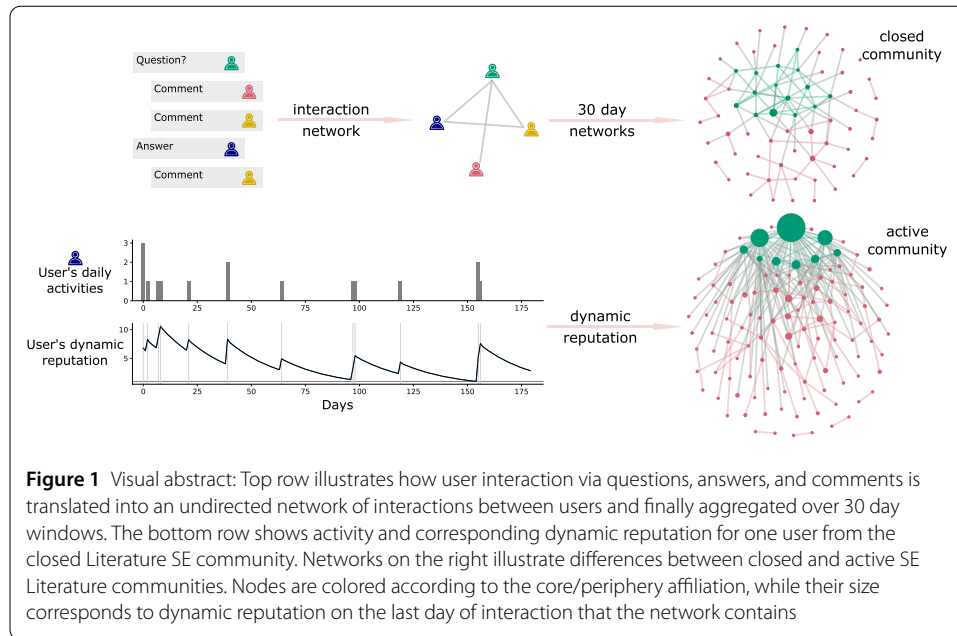
Unlike small offline knowledge-sharing groups, online communities consist of a large number of members where repeatable mutual interactions between all members are not possible. Thus, the trustworthiness of individuals in these communities has to be assessed and signaled using other means. It was shown that the reputation of an individual within the community is a strong signal of her trustworthiness that can override the main sources of social bias [5]. The reputation helps users manage the complexity of the collaborative environment by signaling out trustworthy members.

In the past two decades, we have witnessed the emergence of an online knowledge-sharing community Stack Overflow, which has become one of the most popular sites in the world and the primary knowledge resource for coding. The success of Stack Overflow led to the emergence of similar communities on various topics and formed the Stack Exchange (SE) network.[1] The advancement of Information and communication technologies (ICTs) have enabled faster and easier creation and sharing of knowledge, but also the access to a large amount of data that allowed a detailed study of their emergence and evolution [6], as well as user roles [7], and patterns of their activity [8–10]. However, relatively little attention has been paid to the sustainability of SE communities. Most research focused on the activity and factors that influence the users' activity in these communities. Factors such as the need for experts and the quality of their contributions have been thoroughly investigated [11]. It was shown that the growth of communities and mechanisms that drive it might depend on the topic around which the community was created [12].

In this paper, we investigate the role of network structure and social trust dynamical user reputation in the sustainability of a knowledge-sharing community. Research on the sustainability of social groups shows that social interaction and their structure influence the dynamics and sustainability of social groups [13–16]. Due to large number of users and the smaller probability of repeated interactions dyadic trust between members may not play an essential role in the group dynamics of knowledge-sharing communities. However, it is known that the reputation of users, one of the proxies of trust in online communities, is the primary for them to become and maintain their productive member status [17–19].

With the proliferation of misinformed decisions, it is crucial to understand how to foster communities that promote collaborative knowledge exchange and understand how cooperative norms of trustworthy behavior emerge. The way people interact, specifically the structure of their interactions [20], and how inclusive and trustworthy the key members of the community can influence the sustainability of the knowledge-sharing communities. Although the topic and early adopters are essential in establishing a new SE community, they are not sufficient for sustainability. The current SE network has several examples of communities where the first instance of the community did not survive the SE evaluation process and was shut down, while the second attempt resulted in a sustainable community. Focusing on attempts to establish a community on the same or similar topic with a different outcome allows us to investigate the relevance of social network structure and social trust in the sustainability of knowledge-sharing communities. They are particularly relevant if we wish to understand why some communities established themselves in their second attempt. For those pairs of communities, the topic is the same, and all the initial

---

[1]More information about Stack Overflow is available at: https://stackoverflow.co/ and broad introduction to Stack Exchange (SE) network is available at: https://stackexchange.com/tour. Visit https://area51.stackexchange.com/faq for more details about closed and beta SE communities and the review process.

Figure 1 Visual abstract: Top row illustrates how user interaction via questions, answers, and comments is translated into an undirected network of interactions between users and finally aggregated over 30 day windows. The bottom row shows activity and corresponding dynamic reputation for one user from the closed Literature SE community. Networks on the right illustrate differences between closed and active SE Literature communities. Nodes are colored according to the core/periphery affiliation, while their size corresponds to dynamic reputation on the last day of interaction that the network contains

SE platform requirements were satisfied, but something else was crucial for community decay in the first attempt and its in the second.

Our methods and key results are summarised in a visual abstract in Fig. 1. In our main analysis, we analyze four pairs of SE communities and study the differences in the evolution of social structure and trust between closed and active communities. We have selected four topics from the STEM and humanities: astronomy, physics, economics, and literature. We focus on topics where we could find a matched pair of closed and active communities to control for the differences in topic popularity and, partially, community size. For this reason alone, we do not include Stack Overflow as the most popular community in our analysis. We analyze each pair's early stages of evolution and look at the differences between active and closed communities. Specifically, we map the interactions onto complex networks and examine how their properties evolve during the first 180 days of communities' existence. Using complex network theory [21] we quantify the structure of these networks and compare their evolution in active and closed communities on the same topic. We pay special attention to the core-periphery structure of these networks since it is one of the most prominent features of social networks [22]. We examine how core-periphery structure of active and closed communities evolve and analyze their difference. We show that active communities have a higher value of local normalized clustering and a more stable core membership. On average, the core of the sustainable communities has higher inner connectivity.

To study the evolution of social trust, we adapted the Dynamic Interaction Based Reputation Model (DIBRM) [23]. The model allows us to quantify the trust of each individual over time. We can quantify members' mean and total trust within the core and periphery and follow their evolution through time. The mean reputation of members is higher in sustainable communities than in closed ones, indicating higher levels of social trust. Furthermore, the mean reputation of core members of active communities is constantly above the mean reputation of core members in closed communities, indicating that the creation of trust in the early stages of a community's life may be crucial for its survival.

Our results show that social organization and social trust in the early phases of the life of a knowledge-sharing community play an essential role in its sustainability. Our analysis reveals differences in the evolution of these properties in communities on different topics.

The paper is organized as follows. In Sect. 2 we give a short overview of previous research. Section 3 describes the data and outlines some specific properties of each community. In Sect. 4 we describe the measures and models used for describing the local organization and measuring reputation. Section 5 shows our results. Finally, we discuss our results and selection of model parameters and time window, as well as its consequences in Sect. 6.

2 Previous research

The availability of data from the SE network led to detailed research on the different aspects of dynamics of knowledge sharing communities [6, 8–10], the roles of users [7], and their motivations to join and remain members of these communities [24–28]. The focus of the research in the previous decade was on the evolution of activity in SE communities and the different factors that influence this growth. Ahmed et al. [29] have investigated differences between technical and non-technical communities and showed that within the first four years, technical communities have a higher growth rate, more activity, and are more modular. The comparison of UX community in SE and Reddit [30] showed that the Reddit community grows faster, while SE becomes less diverse and active over time. Special attention was paid to the activities of individual users. In Ref. [31] authors argue that while the overall quality of the answers, measured in the answer score, decays over time, the quality of the answers of the individual user remains constant. This observation suggests that good answerers are *born* and not made within the community. Reputation is used as a proxy for the recognition of experts [32] by other members. However, contrary to common sense, the authors show that the presence of experts can reduce the activity of other members [32]. In [12] authors explore the role of self-and cross excitation in the temporal development of user activity. Differences between growing and declining communities and communities on STEM and humanities topics were explored. Their results show that the early stages of growing communities are characterized by the high cross-excitation of a small fraction of popular users. In contrast, later stages exhibit strong long-term self-excitation in general and cross-excitation by casual users. It was also shown that cross-excitation with power users is more important in the humanities than in STEM communities, where casual users have a more critical role.

A relatively small number of papers focus on the sustainability of SE communities. In Ref. [11], authors examine SE sites through an economic lens. They analyze the relationship between content production based on the number of participants and activities and show that an increase in the number of questions (input) increases the number of answers (output). In their works, Oliveira et al. [33] investigate activity practices and identify the tension between community spirit as proclaimed in SE guidance and individualistic values as in reputation measurement through focus groups and interviews.

Our assumption about the relevance of the structure of social networks in the sustainability of knowledge-sharing communities is supported by research on other social groups. Various factors influence the emergence [34, 35], the evolution, and the sustainability of the groups [13, 20, 36, 37]. The number of committed members [37] and the minimal level of interdependence between members [35] are important factors for the emergence of the

community. The levels of activity have an important role in the emergence and stability of social groups [34, 37], while social factors, such as the size of the group, number of social contacts, or social capital, influence their emergence and collapse [13–16].

Another important branch of research of interest in the sustainability of online communities is the topic of trust. While ICTs make it easier for individuals to establish and maintain social contacts and exchange information and goods, they are also exposed to new risks and vulnerabilities. Social trust relationships, based on positive or negative subjective expectations of another person's future behavior, play an important but largely unexplored role in managing those risks. Recent works show that the vital element of trust is the notion of vulnerability in social relations, and as negative expectations of a trustee's behavior most often imply damage or harm to the trustor, decisions about which users to trust in an online community become paramount [38–40].

In communities such as SE, individuals have three sources of information to rely on when deciding to trust someone in a specific context: (1) knowledge of previous interactions, (2) expectations about future interactions, and (3) indirect information gained through a broader social network. Suppose that the number of active users in such a community increases over a more extended period. In that case, the individuals have little or no history together, no direct interactions, and almost no memory of past interactions. In that case, the social network created by the community becomes a crucial source of information. Therefore, from a network perspective, trust can be the result of reputational concerns and can flow through indirect connections linking actors to one another [40, 41].

In that case, users rely on reputation as a public measure of the reliability of other users active within the same community. Reputation is often quantified based on the history of behavior valued or promoted by a set of community norms and, as such, represents a social resource within the community [42–44]. Since reputation is public information, it is also an incentive. Agents with high reputations are motivated to act trustworthy in the future in order to preserve their status in the community [41]. This idea is supported by psychological findings suggesting that trust is primarily motivated by effects produced by the act of trust itself, regardless of more rational or instrumental outcomes of trustworthy behavior [39].

In terms of modeling collective trust and reputation in online communities, knowledge about past behaviors can be implemented in a trust model in different ways. When estimating trust between agents in a social network, graph-based models focus on the topological information, position, and centrality of agents in a social network to estimate both dyadic and collective measures of social trust. On the other hand, interaction-based models, such as the dynamic reputation model implemented in this paper (DIBRM) [23] estimate trust or reputation based on the frequency and type of agent's interactions over time without taking into account the structure and topology of the interactions between different agents in a network.

3 Data

In our main analysis, we focus on pairs of closed and active SE communities matched by topic. Astronomy, Literature, and Economics are currently active communities. All three communities thrived the second time they were proposed. The first attempt to create communities on these topics resulted in website closure within a year. We add to the comparison the early days of the Physics community and compare its evolution with the closed

Theoretical Physics community. The topics of these communities are not identical, but it is safe to assume that there is a high overlap in user demographics and interests. For these reasons, we treat this pair in the same manner as others. Furthermore, to further solidify our results we have examined the early evolution of four additional communities: Mathematics, Mathematica, Startup Business, Startups. These communities are used to inspect the robustness of our main analysis by comparing main communities with others of similar size, user growth, and activity trends.

The SE data are publicly available and released at regular time intervals. We are primarily interested in the activity and interaction data, which means that we extract the following information for posts (questions and answers) and comments: (1) for each post or comment, we extract its unique ID, the time of its creation, and unique ID of its creator - user; (2) for every question, we extract information about IDs of all answers to that question and ID of the accepted answer; (3) for each post, we collect information about IDs of its related comments. The data contains information about the official SE reputation of each user but only as a single value measuring the final reputation of the user on a day when the data archive was released. Due to this significant shortcoming, we do not include this information in our analysis. In SE, users can give positive or negative votes to questions and answers and mark questions as favorites. However, the data is again provided as a final score recorded at the release. Since this does not allow us to analyze the evolution of scores, we omit this data from our analysis.

All SE communities follow the same path from their creation until they are considered mature enough or closed. In a *Definition* phase, a small number of SE users start by designing a community by proposing hypothetical questions about a certain topic. A successful *Definition* phase is followed by a *Commitment* phase. In this phase, interested users commit to the community to make it more active. The *Beta* phase, which follows after the *Commitment* phase, is the most important. It consists of two steps: a three-week private beta phase, where only committed users may ask/answer/comment questions, and a public beta phase when other members are allowed to join the community. The duration of the public beta phase is not limited. Depending on this analysis, there are three possible outcomes: (1) the community is considered successful and it graduates; (2) the community is active but needs more work to graduate, which means that the public beta phase continues; (3) the community dies and the site is closed. The community evaluation/review process is guided by simple metrics: the average number of questions per day, average number of answers per question, percentage of answered questions, total number of users and number of avid users, and average number of visits per day. However, it should be noted that process is not straightforward and that decision criteria have substantially changed in previous years and sometimes exceptions are made for specific communities.[2]

We study how the social network properties of these social communities and the social trust created among their members evolve during the first 180 days. The first 90 days are recognized as the minimal time a newly established community should spend in the beta phase. We investigate a period that is twice as long since closed communities were active between 180 and 210 days. Given that differences in the first few months of the life of the

---

[2]For example, in 2022 59 websites graduated according to new criteria established in 2019 (which excluded questions per day metric), but as explained in the announcement (https://meta.stackexchange.com/questions/374096/congratulations-to-the-59-sites-that-just-left-beta) exception was made for the AI community which graduated although it didn't meet the criteria that minimum 70% questions have at least one upvoted answer.

Table 1 Community overview for first 180 days according to SE evaluation criteria

Site	Status	Answered	Questions per day	Answer ratio
Physics	Closed	83%	1.93	1.64
	Active	**93**%	**11.76**	**2.74**
Literature	Closed	79%	1.77	1.65
	Active	74%	5.04	1.10
Astronomy	Closed	**95**%	2.62	2.02
	Active	**96**%	3.57	1.49
Economics	Closed	68%	2.04	1.25
	Active	84%	5.66	1.37
Stack Exchange criteria	Excellent	>90%	>10	>2.5
	Needs some work	<80%	< 5	<1

online community can help predict its survival and evolution [45], we focus on the early evolution of SE sites.
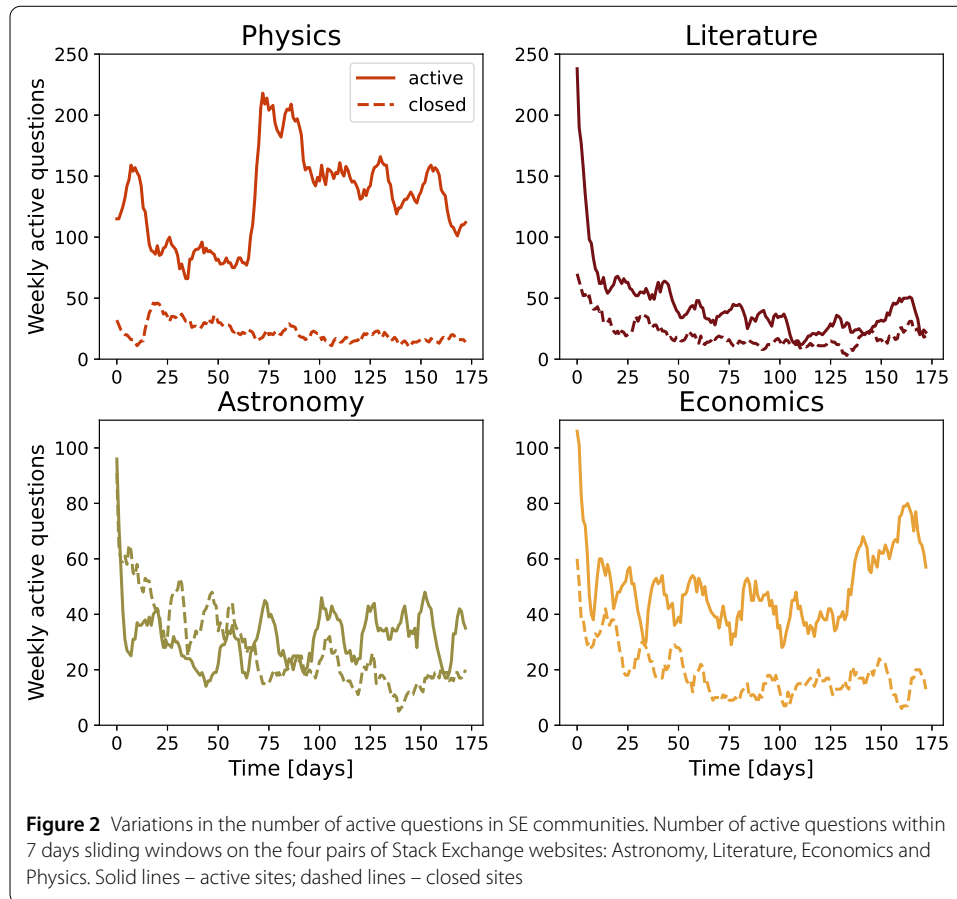
Although the official review of SE communities in the beta phase is mostly based on simple activity indicators such as the number of questions or ratio of answers to questions,[3] these simple metrics do not provide enough information to differentiate between closed communities and those that have been proven to be sustainable in the long term. This may explain why the official guidelines for SE community review have changed and have been applied inconsistently.

Table 1 shows the values of some of these measures at 180 days point for considered communities. Although the Physics community had better metrics than Theoretical Physics and other considered communities, we see that these differences are not as apparent if we compare the remaining three pairs of communities. For instance, some of the parameters for the closed Astronomy community, for example, the percentage of answered questions and answer ratio, were better than for the community that is still active.

Another simple indicator can be the time series of active questions for the 7 days shown in Fig. 2. The question is considered active if it had at least one activity, posted answer, or comment, during the previous 7 days. The four pairs of compared communities show that active communities have a higher number of active questions after 180 days. Although this difference is evident for the Physics and Economics community, Fig. 2 shows that its value is smaller for Astronomy and Literature. Furthermore, in the case of Astronomy, the closed community had a higher number of active questions in the first 75 days.

The values of the measures shown in Tables 1 and A1 in Additional file 1, and Fig. 2 suggest that these simple measures are not good indicators of long-term sustainability. Therefore, we need a deeper understanding of the structure and dynamics of the community to understand the factors behind its sustainability. All communities must start with the same number of interesting questions, the same number of committed users, and satisfy the same thresholds to enter the public beta phase. These basic aggregated statistics are not enough to differentiate between active and closed communities. Hence, other factors determine the sustainability of communities. We investigate the role of social interaction structure and the dynamics of collective trust in the sustainability of SE communities.

---

[3]https://stackoverflow.blog/2011/07/27/does-this-site-have-a-chance-of-succeeding/

Figure 2 Variations in the number of active questions in SE communities. Number of active questions within 7 days sliding windows on the four pairs of Stack Exchange websites: Astronomy, Literature, Economics and Physics. Solid lines – active sites; dashed lines – closed sites

4 Method

We are interested in the position of trustworthy members in SE communities and how active and closed communities differ regarding this factor. First, we map the interaction data onto networks and analyze their properties and how they evolve during the first 180 days. Furthermore, we use the dynamical reputation model to estimate the trustworthiness of each member of the community and the dynamics of collective trust by studying the evolution of the mean value of reputation in the community. The entire analysis was done in Python, and the entire code for reproducing the results and figures is publicly available in an online repository.[4]

4.1 Network mapping

We treat all user interactions, answering questions, posting questions or comments, and accepting answers equally. We construct a network of users where the link between two nodes, users i and j, exists if i answers or comments on the question posted by j and vice versa, or i comments on the answer posted by j and vice versa, i accepts the answer posted by user j. We do not consider the direction or frequency of the interaction between users i and j; thus, the obtained networks are unweighted and undirected.

We create a network snapshot $G(t, t + \tau)$ at the time t for the time window length τ. Two users (i, j) are connected in a network snapshot $G(t, t + \tau)$ if they have had at least one

---

[4]https://github.com/ana-vranic/Stack-Exchange-communities

interaction during the time $[t, t + \tau]$. Our first network accounts for interaction within the first 30 days $G[0, 30)$, and we slide the interaction window by one day and finish with $G[149, 179)$ network. This way, we create 150 interaction networks for each community. By sliding the time window by one day, we create two consecutive networks that overlap significantly. In this way, we can capture subtle structural changes resulting from daily added/removed interactions. We calculate the different structural properties of these networks and analyze how they change over 180 days.

4.2 Clustering

There are many local and global measures of network properties [21]. These measures are not independent. However, it was shown that the degree distribution, degree-degree correlations, and clustering coefficient are sufficient to fully describe most complex networks, including social networks [46]. Furthermore, research on the dynamics of social group growth shows that links between persons' friends who are members of a social group increase the probability that that person will join that social group [47]. Successful social diffusion typically occurs in networks with a high value of the clustering coefficient [48]. These results suggest that higher local cohesion should be a characteristic of sustainable communities.

   The clustering coefficient of a node quantifies the average connectivity between its neighbors and the cohesion of its neighborhood [21]. It is a probability that two neighbours of a node i are also neighbours, and is calculated using the following formula:

$$c_i = \frac{e_i}{\frac{1}{2}k_i(k_i - 1)} \; . \tag{1}$$

Here e_i is the number of links between the neighbours of the node i, while $\frac{1}{2}k_i(k_i - 1)$ is the maximum possible number of links determined by the degree of the node k_i. The clustering coefficient of the network C is the value of the clustering averaged over all nodes. We investigate how the clustering coefficient in an SE community changes over time by calculating its value for all network snapshots. We normalize the clustering coefficients with the value of expected clustering for the random Erdos-Renyi network with the same number of nodes N and links L: $c_{er} = p = \frac{2L}{(N(N-1))}$ [21, 49]. We compare normalized clustering coefficient for active and closed communities on the same topic to better understand the evolution of cohesion of these communities.

4.3 Core-periphery structure

Real networks, including social networks, have a distinct mesoscopic structure [22, 50]. The mesoscopic structure is manifested either through the community structure or the core-periphery structure. Networks with a community structure consist of a certain number of groups of nodes that are densely connected, with sparse connections between groups. Networks with core-periphery structures consist of two groups of nodes, with higher edge density within one group, core, and between groups. However, low edge density in the second group, periphery [22]. Research on user interaction dynamics in SE communities shows that there is a small group of highly active members who have frequent interactions with casual or low active members [8, 12]. These results indicate that we should expect a core-periphery structure in SE communities. The classification of nodes

into one of these two groups provides information on their functional and dynamic roles in the network.

To investigate the core-periphery structure of SE communities and how it evolves over time, we analyze the core-periphery structure of every network snapshot. For this purpose, we use the Stochastic Block Model (SBM) adapted for the inference of the core-periphery of the network structure [22].

SBM is a model where each node belongs to one group in the given network *G*. For the core-periphery structure, the number of blocks is two. Thus, the elements of the vector θ_i are 1 if the node *i* belongs to the core or 2 for the periphery. The block connectivity matrix $\{\boldsymbol{p}\}_{2x2}$ specifies the probability p_{rs} that nodes from group *r* are connected to nodes in group *s*, where $r, s \in \{1, 2\}$.

The SBM model seeks the most probable model that can reproduce a given network G. The probability of having model parameters θ, $\boldsymbol{p}$ given network *G* is proportional to the likelihood of generating network *G*, $P(G|\theta, \boldsymbol{p})$, prior on SBM matrix $P(\boldsymbol{p})$ and prior on block assignments $P(\theta)$:

$$P(\theta, \boldsymbol{p}|G) = P(G|\boldsymbol{p}, \theta)P(\boldsymbol{p})P(\theta), \tag{2}$$

The likelihood of generating a network *G* is defined as:

$$P(G|\theta, \boldsymbol{p}) = \prod_{i<j} p_{r_i s_j}^{A_{ij}}(1 - p_{r_i s_j})^{1-A_{ij}}, \tag{3}$$

where the adjacency matrix element A_{ij} is equal to 1 whenever nodes *i* and *j* are connected and it is 0 otherwise.

Prior on $\boldsymbol{p}$ is the uniform distribution over all block matrices whose elements satisfy the constraint for the core-periphery structure $0 < p_{22} < p_{12} < p11 < 1$. Prior on θ consists of three parts: the probability of having 2 blocks; given the number of blocks, probability $P(n|2)$ of having groups of sizes $\{n_1, n_2\}$ and probability $P(\theta|n)$ of having particular assignments of nodes to blocks.

To fit the model, we follow the procedure set by the authors of Ref. [22] and use the Metropolis-within-Gibbs algorithm. For each 30 days snapshot network, we run 50 iterations and choose the model parameters θ and p according to the minimum description length (MDL). MDL does not change much among inferred core-periphery structures, see Fig. A1 in Additional file 1, while looking into the Adjusted Rand Index (ARI), we can notice that difference exists. Still, the ARI between pair-wise compared partitions is significant (ARI > 0.9), indicating the stability of the inferred structures. The definition and detailed descriptions of MDL and ARI are given in the Additional file 1.

4.4 Dynamic reputation model

Any dynamical trust or reputation model has to take into account distinct social and psychological attributes of these phenomena in order to estimate the value of any given trust metric [43]. First, the dynamics of trust are asymmetric, meaning that trust is easier to lose than to gain. As part of asymmetric dynamics, to make trust easier to lose, the trust metric has to be sensitive to new experiences, recent activity, or the absence of the user's activity while still maintaining the non-trivial influence of old behavior. The impact of

new experiences must be independent of the total number of recorded or accumulated past interactions, making high levels of trust easy to lose. Finally, the trust metric must detect and penalize behavior that deviates from community norms.

We estimate the dynamic reputation of SE users using the Dynamic Interaction Based Reputation Model (DIBRM) [23]. This model is based on the idea of dynamic reputation, which means that the reputation of users within the community changes continuously over time: it should rapidly decrease when there is no registered activity from the specific user in the community, reputation decay, and it should grow when frequent, constant interactions and contributions to the community are detected. The highest growth in users' reputations is found through bursts of activity followed by a short period of inactivity.

Our model implementation does not distinguish between positive and negative interactions in SE communities. Therefore, we treat any interaction in the community, posting a question, answer, or comment, as a potentially valuable contribution. The evaluation criteria for SE websites that go through beta testing described in Additional file 1 do not distinguish between positive and negative interactions. The percentage of negative interactions in the communities we investigated was below 5%, see Table A2 in Additional file 1. Filtering positive interactions would also require filtering out comments because the community does not rate them. That would eliminate a large portion of direct interactions between community users, which is essential for estimating their reputation. The only negative aspect of behavior in our model is the absence of valuable contributions - the user's inactivity. This behavior can be seen as a deviation from community norms as we look at new communities in the early stages of development, where constant contributions are crucial to community growth and survival.

In DIBRM, the reputation value for each user of the community is estimated by combining two different factors: (1) *reputation growth* - the cumulative factor that represents the importance of users' activities; (2) *reputation decay* - the forgetting factor that represents the continuous decrease in reputation due to inactivity. In the case of SE communities, the forgetting factor has a literal meaning, as we can assume that active users forget users' past contributions as their attention is captured by more recent content.

In the bottom left part of Fig. 1 we see an example of reputation dynamics for a single user. There are bursts of reputation growth after multiple interactions are recorded, like in the case of two interactions in a single day recorded between days 25 and 50, followed by a period of inactivity which leads to reputation decay. In this case, the decay is interrupted by a single recorded activity before the 75th day, but then an even longer inactivity period ensued, leading to a decay that reduced the reputation of the user nearly to 0 before the 100th day. Two contrasting examples of real user reputation are explained in the Additional file 1 (Fig. A2).

Reputation dynamics revolves around the varying influence of past and recent behavior. Thus, DIBRM has two components: *cumulative factor* - estimating the contribution of the most recent activities to the overall reputation of the user; *forgetting factor* - estimating the weight of past behavior. Estimating the value of recent behavior starts with the definition of the parameter storing the basic value of a single interaction I_{b_n}. The cumulative factor I_{c_n} then captures the additive effect of successive recent interactions. In Fig. 1 we see this cumulative effect with two consecutive interactions (gray vertical lines) after day 150 which sudden jump in reputation previously reduced to zero. The reputational contribution I_n of the most recent interaction n of any given user is estimated in the following

way:

$$I_n = I_{b_n} + I_{c_n} = I_{b_n} \left(1 + \alpha \left(1 - \frac{1}{S_n + 1} \right) \right). \tag{4}$$

Here, α is the weight of the cumulative part, and S_n is the number of sequential activities. If there is no interaction at t_n, this part of interactions has a value of 0. An essential property of this component of dynamic reputation is the notion of sequential activities. Two subsequent interactions by a user are considered sequential if the time between these two activities is less than or equal to the time parameter t_a that represents the time window of interaction. This time window represents the maximum time spent by the user to make a meaningful contribution, post a question or answer, or leave a comment,

$$\Delta_n = \frac{t_n - t_{n-1}}{t_a} \ . \tag{5}$$

If $\Delta_n < 1$, the number of sequential activities S_n will increase by one, which means that the user continues to communicate frequently. However, large values Δ_n significantly increase the effect of the forgetting factor. This factor plays a vital role in updating the total dynamic reputation of a user at each time step, after every recorded interaction:

$$T_n = T_{n-1} \beta^{\Delta_n} + I_n \ . \tag{6}$$

Here, β is the forgetting factor. In our model implementation, the trust is updated each day for every user regardless of their activity status. Therefore, the decay itself is a combination of β and Δ_n: the more days pass without recorded interaction from a specific user, the more their reputation decays. Lower values of β lead to faster trust decay, as shown in Fig. A2 in the Additional file 1. In Fig. 1 we observe this long-tailed reputation loss when the user has more than 25 inactive days between days 120 and 150, reducing the reputation almost to 0.

For this work, we select the following values of these parameters: (1) we set the basic reputation contribution $I_{bn} = 1$, which means that each activity contributes 1 to the dynamical reputation; (2) for the cumulative factor α we choose the value 2 and place higher weight on recent successive interactions; (3) forgetting factor β we select the value 0.96; 4) the value of $t_a = 2$. By setting $\alpha > 1$ we enable faster growth of reputation due to a large number of subsequent interactions; see Fig. A2 in Additional file 1. Furthermore, by setting the value of $\beta < 1.0$, we increase the penalty for long inactivity periods; see Fig. A2 in Additional file 1. We discuss the selection of model parameters and their consequences in detail below. The selected values of parameters are used to measure the dynamical reputation of users in all four pair SE communities. Given these parameter values, the minimal reputation of the user immediately after having made an interaction in the SE community is 1. This reputation will decay below 1 if the user does not perform another interaction within the one-day window. Users with a reputation below the value 1 are considered inactive and *invisible* in the community; that is, their past contributions at that time are unlikely to impact other users.

4.4.1 The choice of model parameters

In this work, we used snapshots of the network of 30 days. This period corresponds to the average month, and it is common in the analyses of the structure and dynamics of social networks [51–53]. Still, there is no well-specified procedure to choose the time window. Previous studies have shown that if τ is small, subnetworks become sparse, while for too large sliding windows, some important structural changes cannot be observed [52, 54]. Thus, we have analysed how the time window choice influences our results. Figure A11 in Additional file 1 shows how considered network properties and dynamical reputation depend on the time window size for active and closed communities in case of Astronomy communities. We observe that fluctuations of all measures are more pronounced for a time window of 10 days than for 30 and 60 days. However, we find that while the structural properties of networks evolve at different rates over varied time windows, the trends remain very similar. The qualitative difference observed between closed and active communities is independent of the time window size, especially when comparing the 30 and 60 day windows. The 30-day time window ensures enough interaction, even for closed communities, while the number of observation points remains relatively high. For these reasons, we choose a sliding window of 30 days.

The initial purpose of DIBRM was to replicate the dynamics of the official SE reputation metric [23, 55]. In previous studies [55] the official SE reputation is obtained with $t_a = 2$, $\alpha = 1.4$, $\beta = 1$. This configuration of model parameters implies that there is no reputation decay and points toward the fact that the official SE reputation is hard to lose. Our application is oriented towards estimating a reputation metric which takes into account the fundamental properties of social trust, i.e. reputation decreases with members' inactivity, so we opted for a different set of parameter values.

For the basic reputation contribution of a single interaction, we selected $I_{bn} = 1$, and, at the same time, this is the threshold value of an active user. This value is intuitive as every interaction has the initial contribution of +1 to the user's reputation, although the previous works have used values of +2 and +4. Following the previous work and after examining the median/average time between subsequent interactions of the same user, we selected $t_a = 1$, which also means that the reputation in our model will be updated every day during the time window of the analysis, regardless of whether the user is active or not.

The combination of parameters α and β can significantly influence the dynamic of the single user reputation, as shown in Fig A2. We show that higher values for parameter $\alpha = 2$, highlight the burst of user activity and frequent interaction. On the other hand, the parameter beta is the forgetting factor, which at the same time determines the weight of past interactions and the reputational punishment due to user inactivity. Here, we need to select the parameter β value, so we include forgetting due to inactivity but do not penalize it too much. In Fig. A2, we show how different values of parameter β influence the time needed for a user's reputation to fall on value $I_n = 1$ due to the user's inactivity and value of dynamical reputation at the moment of the last activity. The higher value of the parameter β and the initial dynamical reputation of the users, the longer it takes for the user's reputation to fall to the baseline value. For parameters $\beta = 0.9$ and $I_n = 5$, the user's reputation drops to value $I_n = 1$ after less than 20 days, while this time is doubled for $\beta = 0.96$. We see that for higher values of the parameter β, the time it takes for I_n to drop to 1 becomes longer and that the initial value of the reputation becomes less important.
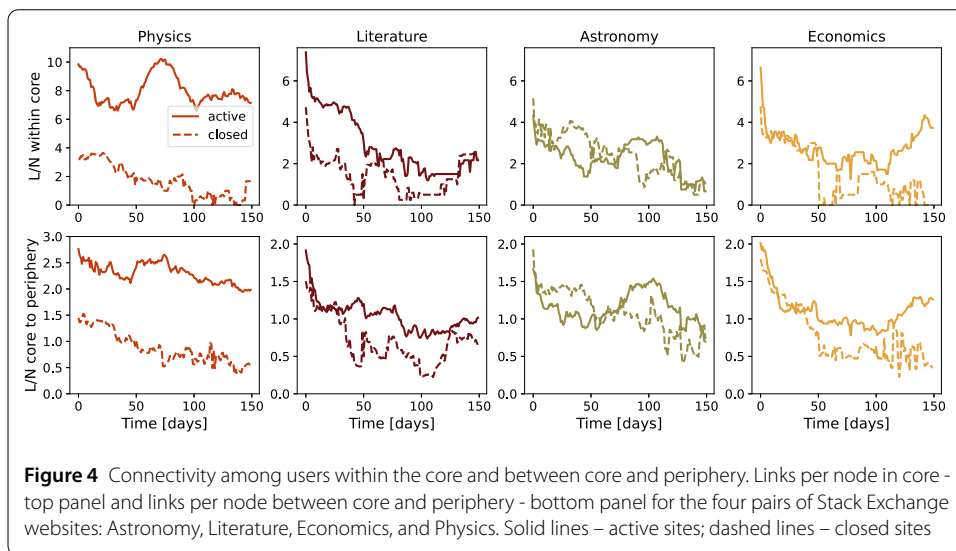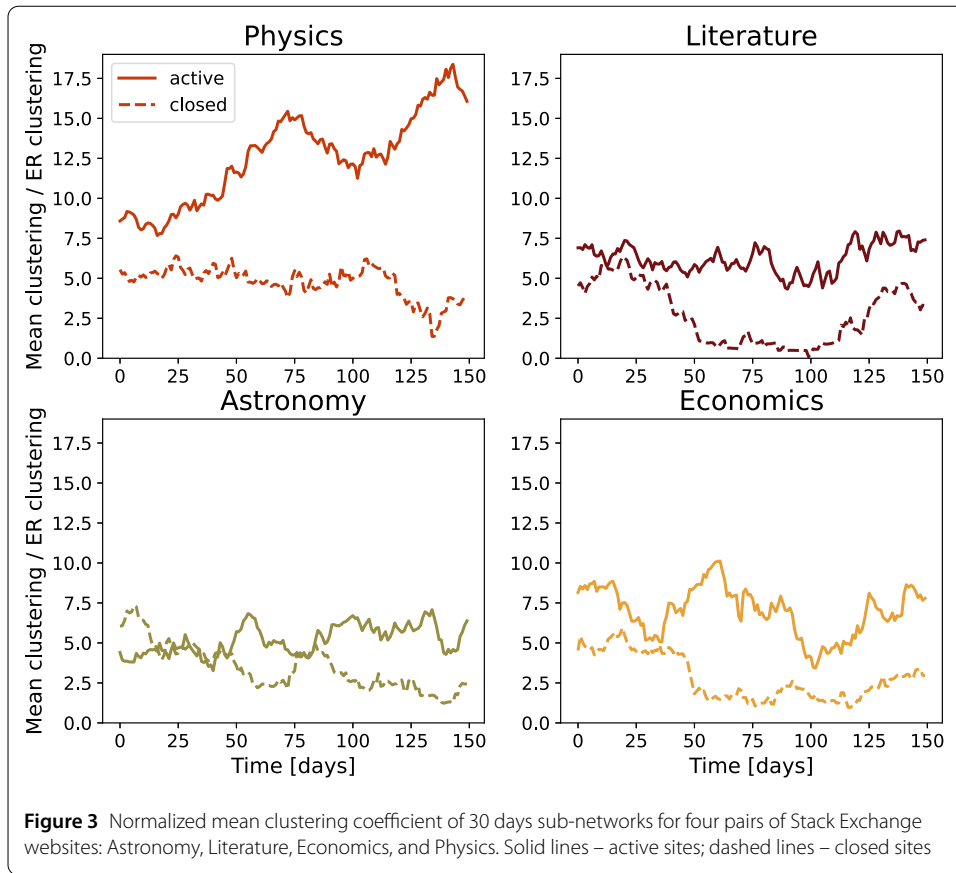
We estimated the difference between the number of users who had at least one activity in the 30-day window and the number of users with a reputation greater than 1 during the same period for different parameter β values. We calculated the root mean square error (RMSE) between the time series of the number of active users for τ = 30 and different values of β parameters; see Fig. A12 in Additional file 1. The minimal difference between these two variables is for β between 0.94 and 0.96 for both active and closed communities. Since we want to compare communities, we select β = 0.96. Our analysis reveals that the reputational decay parameter β set at 0.96 does not reduce the number of active users (based on their dynamic reputation) below the actual number of users who have been active (interacted with the community) in the time window of 30 days; see Fig. A13 in Additional file 1. Furthermore, we examine and compare the trends of two types of time series: (1) time series of active users, according to dynamical reputation; (2) time series of permanent users, users who were active in a given sliding window and continued to be active in the next one. Figure A14 in Additional file 1 shows that while the absolute number of users differs in these time series, they follow similar trends for all communities.

5 Results

5.1 Clustering and core-periphery structure of knowledge-sharing networks

We first analyze the structural properties of SE communities and examine the difference between active and closed ones. We calculate the normalized mean clustering coefficient for 30-day window networks and examine how it changes over time. Figure 3 shows the evolution of the normalized mean clustering coefficient for the eight communities. All communities that are still active are clustered, with the value of normalized clustering coefficient above 5, with Physics, the only launched community, having the highest value of normalized clustering coefficient during the first 180 days. During the larger part of the observed period, an active community's normalized clustering coefficient is higher than the normalized clustering coefficient of its closed pair. For pairs where active communities are still in the beta phase, some of closed communities have a higher value of the normalized clustering coefficient in the first 50 days. After this period, active communities have higher values of the normalized clustering coefficient. These results suggest that all communities have relatively high local cohesiveness compared to random graphs, however, the value of normalized clustering below the value 5 in the later phase of community life may indicate its decline.

Furthermore, we examine the core-periphery structure of these communities and their evolution. Specifically, we are interested in the evolution of connectivity in the core. Figure 4 shows the change in the number of links between nodes, averaged on the core nodes, $\frac{L_c}{N_c}$ over time. $\frac{2L_c}{N_c}$ is the average degree of the node in the core and, thus, $\frac{L_c}{N_c}$ is the half of the average degree. Again, the Physics community has a much higher value of this quantity than Theoretical Physics during the observed period, indicating higher connectivity between core members. Higher connectivity between core members in the active community is also characteristic of Literature. However, this quantity has the same value for active and closed communities at the end of the observation period. The differences between active and closed communities are not that prominent for Economics and Astronomy, see Fig. 4. Active and closed Economics communities have similar connectivity in the core during the first 50 days. After this period, the connectivity in the core of the active community is twice as large as in the closed community, and the difference grows at

Figure 3 Normalized mean clustering coefficient of 30 days sub-networks for four pairs of Stack Exchange websites: Astronomy, Literature, Economics, and Physics. Solid lines – active sites; dashed lines – closed sites

Figure 4 Connectivity among users within the core and between core and periphery. Links per node in core - top panel and links per node between core and periphery - bottom panel for the four pairs of Stack Exchange websites: Astronomy, Literature, Economics, and Physics. Solid lines – active sites; dashed lines – closed sites

the end of the observation period. The connectivity in the core of the closed Astronomy community is higher than the connectivity in the core of the active community during the first 50 days. However, as time progresses, this difference changes in favor of the active community, while this difference disappears at the end of the observation period.

The difference between active and closed communities is observed compared to the average number of core-periphery edges per network node. The connectivity between core
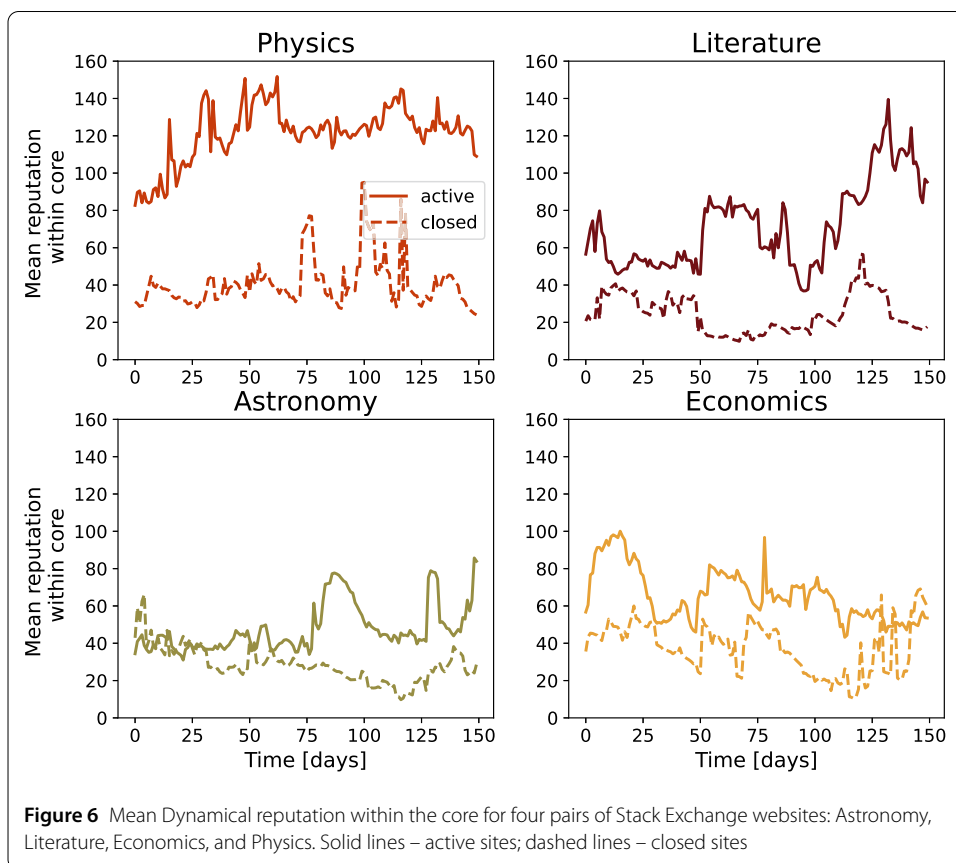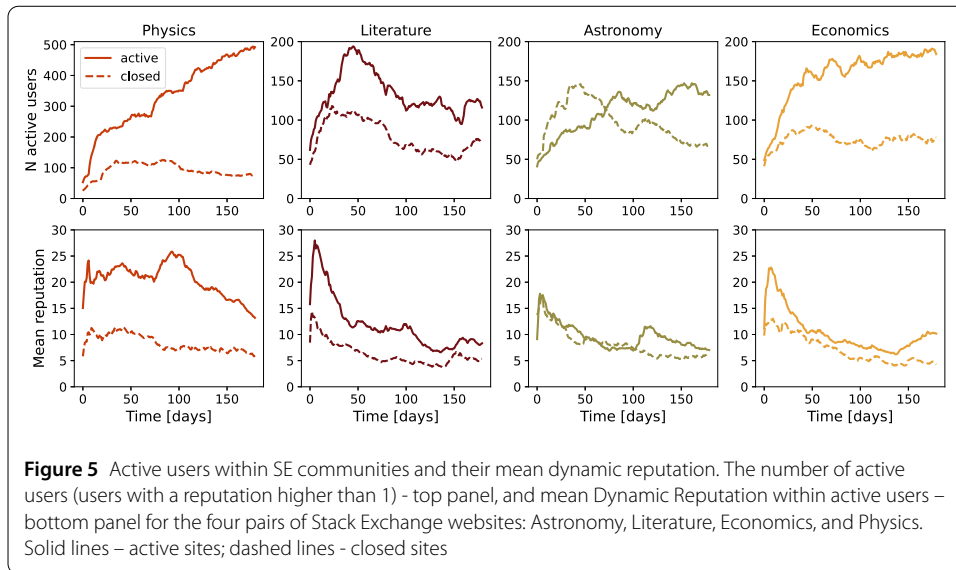
and periphery is higher for the active communities than for the closed ones, see Fig. 4, which is very obvious if we compare Physics and Theoretical Physics communities. Moreover, the Physics community has the highest connectivity compared to all other communities. Active Literature and Economics communities have the same core-periphery connectivity as their closed counterpart. The core of the active Astronomy community has weaker connections with the periphery than the closed community during the first 50 days, see Fig. 4.

Our motivation to examine the core-periphery structure comes from reference [12]. The authors have selected 10% of the most active users and examined their mutual connectivity and connectivity with the remaining users. The split of 10% to 90% users according to their activity may appear arbitrary. The core-periphery provides a more consistent network division based on its structure. However, the connectivity patterns between popular-popular and popular-casual users, shown in Fig. A3 in Additional file 1, are similar to one observed for core-periphery in Fig. 4.

On average, the cores of active communities have a higher number of nodes than closed communities. However, the size of the core relative to the size of the network is similar for active and closed communities (Fig. A4 in Additional file 1). The size of the core fluctuates over time for active and closed communities. The core membership also changes over time. This core membership is changing more for the closed communities. We quantify this by calculating the Jaccard index between the cores of the subnetworks at the moment t_i and t_j. Figure A5 in Additional file 1 shows the value of the Jaccard index between any pair of the 150 subnetworks. The highest value of the Jaccard index is around the diagonal and has a value close to 1. The compared subnetworks are for consecutive days and have a similar structure. The value of the Jaccard index decreases with the number of days between two subnetworks $|t_i - t_j|$ faster in closed communities; see Fig. A6 in Additional file 1. This difference is the most prominent for the Literature communities, while this difference is practically non-existent for Astronomy. The relatively high value of overlap between cores of distant subnetworks for active communities further confirms that the core is more stable in these communities that in their closed counterparts.

5.2 Dynamic reputation of users within the network of interactions

To explore the differences between active and closed communities, we focus on dynamical reputation, our proxy for collective trust in these communities. The number of active users (top panel) and the mean user reputation (bottom panel) for different SE communities are shown in Fig. 5. Except in the case of Astronomy, closed communities generated less engaged users from the start and the number of active users saturated at lower values. In the case of Astronomy, the closed community started with a faster-increasing number of active users. However, within the first two months, their number dropped, while the second time around, the community started slower but kept engaging more users. Only in the still active Physics community is the number of active users an increasing function over the whole 180 day period we have observed. Panels in the bottom show mean reputation among active users, and we see that most of the time, it was higher in the still active communities than in the closed ones. The Physics community kept these mean values more stable at higher levels, whereas in other communities, we note that the initial high mean reputation decays faster. Astronomy is an exciting exception again, where we see a second sudden increase in mean user reputation, which signals an increase in user activity.

Figure 5 Active users within SE communities and their mean dynamic reputation. The number of active users (users with a reputation higher than 1) - top panel, and mean Dynamic Reputation within active users – bottom panel for the four pairs of Stack Exchange websites: Astronomy, Literature, Economics, and Physics. Solid lines – active sites; dashed lines - closed sites

Figure 6 Mean Dynamical reputation within the core for four pairs of Stack Exchange websites: Astronomy, Literature, Economics, and Physics. Solid lines – active sites; dashed lines – closed sites

In addition, we investigate whether and how the core-periphery structure is related to collective trust in the network. Figure 6 shows the mean dynamical reputation in the core of active and closed communities and its evolution during the observation period. There are apparent differences between active and closed communities regarding dynamical reputation. The mean dynamical reputation of core users is always higher in active commu-

nities than in closed. The most significant difference is observed between the Physics and Theoretical Physics communities. The difference between active communities, which are still in the beta phase, and their closed counterparts is not as prominent. However, the active communities have a higher mean dynamical reputation, especially in the later phase of the observation period. The only difference in the pattern is observed for Astronomy communities at the early stage of their life. The closed community has a higher value of dynamic reputation than the active community. This observation is in line with similar patterns in the evolution of mean clustering, core-periphery structure, and mean reputation.

By definition, the core consists of very active individuals. Thus we expect a higher total dynamical reputation of users in the core than the total reputation of users belonging to subnetworks periphery. Figure A7 in Additional file 1 shows the ratio between the total reputation of the core and periphery for closed and active communities and their evolution. The ratio between the total reputation of core and periphery in Physics is always higher than in the Theoretical Physics community. A similar pattern can be observed for Literature communities, although the difference is not as prominent as in the case of Physics. The ratio of total dynamical reputation between core and periphery was higher in the closed Economics community during the early days of its existence. However, this ratio becomes higher for active communities in the later stage of their lives. Communities around the astronomy topic deviate from this pattern, which shows the specificity of these two communities.

To complete the description of the evolution of dynamic reputation, we examine the evolution of the Gini index of dynamical reputation among the active members of SE sites, shown in Fig. A8 in Additional file 1. Both closed and active communities have high values of the Gini index, indicating that the dynamic reputation is distributed unequally among users. Notably, all communities have the highest Gini index at the start, signaling that the inequality in users' activity at the start, and thus their dynamic reputation is the highest. After this initial peak, the Gini index decreases, but it persists at higher levels in communities that are still active than in the closed ones, except in the case of the Astronomy community. In this case, the active community had a higher Gini index until just before the observation period, when the Gini coefficient increased in the closed community.

Figure A9 in Additional file 1 shows the evolution of the assortativity coefficient for users' dynamical reputation. The observed networks are disassortative during the most significant part of 180 days period. Users with high dynamical reputations tend to connect with users with a low value of dynamical reputation in all eight communities. We also compare the degree and betweenness centrality of the users and their dynamical reputation by calculating the correlation coefficient between these measures for each sliding window, see Fig. A10 and detailed explanation in Additional file 1. The correlation between these centrality measures and dynamical reputation is very high. In active communities on physics, economics, and literature topics, the correlation between centrality measures and users' reputation is exceptionally high, above 0.85, and does not fluctuate much during the observation period. There is a clear difference between active and closed communities for these three pairs. The Astronomy pair deviates from this pattern for the first 100 days. After this period, the pattern is similar to one observed for the other three pairs of communities. The results reveal that degree and betweenness centrality are correlated more with a reputation in active than in closed communities.

6 Discussion and conclusions

In this work, we have explored whether the structure and dynamics of social interactions determine the sustainability of knowledge-sharing communities. We have adopted a model of dynamical reputation to measure the collective trust of members and analyzed its dynamics. For this purpose, we use the data from the SE platform of knowledge-sharing communities where members ask and answer questions on focused topics. We selected four pairs of active and closed communities on the same or similar topic. Specifically, two topics are from the STEM field, physics, and astronomy, and two are from social sciences and humanities, economics and literature.

We have examined the evolution of the normalized average clustering coefficient in closed and active SE communities. Our results show that active communities have significantly higher values of clustering coefficient compared to ER graphs of the same size in the later phase of community life than closed communities. In the early phase of communities' lives, the clear difference between active and closed communities is observed only for the physics topic; see Fig. 3. The high value of the normalized clustering coefficient observed for the active Physics community suggests that communities with high local cohesiveness are sustainable and mature faster than others.

The core in active communities is more strongly connected with the periphery than in closed communities, indicating that active members engage more often with occasionally active members; see Fig. 4. These results suggest that active communities are more inclusive than closed ones. Furthermore, our analysis shows that average connectivity between core members is not as crucial to community sustainability as expected. Although active Physics and Economics communities exhibit much higher connectivity in the core than their closed counterparts, this is not true for communities focused on astronomy and literature. However, our results show that a member's lifetime in the core is longer for active communities, indicating a more stable core in active communities.

Analysis of the evolution of the core-periphery and its connectivity patterns suggests a higher trust between active and sporadically active members. To further explore this, we have adapted the dynamical reputation model [23], which allowed us to follow the evolution of trust of each member.

The total dynamical reputation of core members during their first 180 days was higher for active communities than for their closed counterparts. While relative core size is less than 40%, Fig. A4 in Additional file 1, the ratio between the total reputation of nodes in the core and ones in the periphery is consistently above 0.5, indicating that the average reputation of members in the core is higher than the reputation of the node in the periphery. The ratio between the total reputation of core and periphery nodes has a higher value in the active community of Physics, Literature, and Economics. For most of the 180 days, this ratio has a value higher than one. The Astronomy communities are outliers, but the core members have a higher total reputation than members on the periphery, even for these two communities. Our results imply that the most trusted members in the community are the core members, who also generate more trust in active communities. They have a higher reputation generated through interactions with both core and nodes in the periphery, see Fig. 6. Furthermore, the overall levels of trust are higher in active communities, which is reflected in the fact that the mean user reputation is higher in these communities; see Fig. 5.

The choice of the topics and selection of SE communities of a various number of users, question, answer and comments, see Table A1 in the Additional file 1, guarantees, up to a certain extent, the generality of our results. However, there are certain limitations to the generalizability of our findings. While SE communities provide very detailed data that enable the study of the structure and dynamics of knowledge-sharing communities, we must not ignore the fact that they have some properties that make them specific.

SE communities are about specific topics; they mostly bring together people who are passionate about or are experts in a specific field. These communities attract people from the general population. Since we were interested in excluding the factor of the topic in our research, we studied and compared active and closed communities on the same topic. In the SE network, these pairs of communities are pretty rare, which has substantially limited our sample size, leaving the possibility for the occurrence of outliers that do not follow our general conclusions.

To further solidify our results, we have examined the early evolution of four additional communities: Mathematics, Mathematica, Startup Business, and Startups. Mathematics and Mathematica communities graduated early in the process, while both communities on startup topics were closed after spending some time in the public beta phase. Figures A15 and A16 in the Additional file 1 show that both communities on the subject of mathematics exhibit a similar evolutionary path as the Physics community. They have a high mean reputation, stable and relatively large cores with high average trustworthiness of core members, see Fig. A15 in Additional file 1. While the numbers of active users in these two communities and the Physics community differ, we see that this does not influence the average reputation of users or the size of the core. This is even more evident if we compare the Physics community with the closed Startup Business community. We see from Fig. A16 in Additional file 1 that the number of active users grows much faster for this community than for Physics. However, the average reputation in the community is comparable with the ones that were eventually closed, Theoretical Physics and Startups. Furthermore, the core size is comparable with the core of Physics, but the average trustworthiness of core members is similar to one for closed communities. These results demonstrate that even the communities with high early activity and a number of active users will not become sustainable if they do not develop a core of trustworthy members. Startups community has a behavior very similar to Theoretical Physics community. The comparison between two startup communities, shows that despite their difference in the activity levels these communities have similar evolution path during the first 180 days.

We have also decided to map interactions to networks so that the resulting network is unweighted and undirected. We use unweighted edges for a finer distinction between the structure and community dynamics. The number of repeated user interactions is captured with dynamic reputation, while the edges carry only structural information without the number of repeated interactions. Furthermore, as we map interactions to networks using sliding windows, the repeated presence of an edge throughout different windows gives us partial information about the durability and the frequency of the dyadic relationship. Similarly, we opted against directed weights as we are not interested in diffusion or flow of information and undirected edges represent a more parsimonious view of the community structure. However, these choices did have consequences in the choice of core-periphery detection method, and it is possible that with different network mapping, other methods would prove more suitable.

Finally, there are many ways to measure collective trust and reputation in online social communities. We have selected the dynamical reputation model because it was developed to measure reputation in SE communities. Furthermore, the model allowed us to study the evolution of trust in communities. However, the model requires fine-tuning of its parameters and does not distinguish positive from negative interactions. We have selected our parameters to replicate the activity of the SE communities in the time window of τ = 30 days. Our analysis shows that while the choice of the sliding window, τ, may seem arbitrary, the different values do not influence the general conclusions; see Fig. A11 in Additional file 1. The interactions in SE communities are mostly not emotional, and thus, the model is suitable for measuring collective trust in these communities. However, the interaction in other knowledge-sharing communities can be much more emotional, and therefore the dynamical reputation model needs to be adapted to measure reputation in these communities.

Our results show that the trustworthiness of core members thus represents one of the essential parameters for determining community sustainability. Sustainable communities have a core of trustworthy members. The core of sustainable communities is more densely connected, and its connectivity with the periphery is more significant than in closed communities. The observed feature is especially prominent in the Physics community, which is the only active community considered to be mature. As we stated, active communities on topics of astronomy, economics and literature were in the beta phase. However, since December 2021,[5] these communities graduated. The core of sustainable communities exhibits higher degrees of stability during their first 180 days. Sustainable communities have higher local cohesiveness, which is reflected in the relatively high value of the normalized clustering coefficient. Our results show that these conclusions hold for both STEM and humanities topics. However, we do not observe apparent differences between active and closed Astronomy communities for some quantities. In the case of Astronomy and sometimes Economics, we find that closed communities had higher normalized clustering coefficients and higher core-core and core-periphery connectivity during the early phase of community life. These observations suggest that the properties of the network during the early phase of the community's existence may lead to wrong conclusions about its sustainability. Our results also imply that information about community sustainability is hidden in the evolution of different network and trust properties.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-023-00381-x.

Additional file 1. The file contains all additional figures, tables and descriptions regarding the analysis performed in the manuscript. The file is in pdf format. (PDF 3.6 MB)

---

[5]https://stackoverflow.blog/2021/12/16/congratulations-are-in-order-these-sites-are-leaving-beta/

Abbreviations

ARI, Adjusted Rand Index; DIBRM, Dynamic Interaction Based Reputation Model; ICT, Information and communication technologies; MDL, Minimum Description Length; RMSE, Root mean square error; SBM, Stochastic Block Model; SE, Stack Exchange.

Availability of data and materials

The Stack Exchange data can be downloaded from Stack Exchange Data Dump, https://archive.org/details/stackexchange. Area 51 Stack Exchange communities can be downloaded from https://area51.stackexchange.com/. The source code and the datasets generated and analysed during the current study are publicly available at https://github.com/ana-vranic/Stack-Exchange-communities.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author contribution

AV, AT, AA, MMD designed the research. AV, AT and AA collected the data and performed data analysis. All authors wrote and edited the final manuscript. All authors read and approved the final manuscript.

Author details

[1]Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, Belgrade, Serbia. [2]Department of Sociology, Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia. [3]Two desperados, Belgrade, Serbia.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Leydesdorff L (2001) In: A sociological theory of communication: the self-organization of the knowledge-based society. Universal-Publishers, USA. https://doi.org/10.1108/jd.2002.58.1.106.2
2. Leydesdorff L (2012) The triple helix, quadruple helix,…, and an n-tuple of helices: explanatory models for analyzing the knowledge-based economy? J Knowl Econ 3(1):25–35. https://doi.org/10.1007/s13132-011-0049-4
3. Lipkova H, Landová H, Jarolímková A (2017) Information literacy vis-a-vis epidemic of distrust. In: European conference on information literacy. Springer, Berlin, pp 833–843
4. Lucassen T, Schraagen JM (2012) Propensity to trust and the influence of source and medium cues in credibility evaluation. J Inf Sci 38(6):566–577
5. Abrahao B, Parigi P, Gupta A, Cook KS (2017) Reputation offsets trust judgments based on social biases among airbnb users. Proc Natl Acad Sci 114(37):9848–9853
6. Dankulov MM, Melnik R, Tadić B (2015) The dynamics of meaningful social interactions and the emergence of collective knowledge. Sci Rep 5(1):1–10. https://doi.org/10.1038/srep12197
7. Saxena A, Reddy H (2021) Users roles identification on online crowdsourced q&a platforms and encyclopedias: a survey. J Comput Soc Sci 1–33. https://doi.org/10.1007/s42001-021-00125-9
8. Santos T, Walk S, Kern R, Strohmaier M, Helic D (2019) Activity archetypes in question-and-answer (q8a) websites—a study of 50 stack exchange instances. ACM Trans Soc Comput 2(1):1–23. https://doi.org/10.1145/3301612
9. Slag R, de Waard M, Bacchelli A (2015) One-day flies on stackoverflow-why the vast majority of stackoverflow users only posts once. In: 2015 IEEE/ACM 12th working conference on mining software repositories. IEEE, pp 458–461. https://doi.org/10.1109/MSR.2015.63
10. Chhabra A, Iyengar SRS (2020) Activity-selection behavior of users in stackexchange websites. In: Companion proceedings of the web conference 2020, pp 105–106. https://doi.org/10.1145/3366424.3382720
11. Dev H, Geigle C, Hu Q, Zheng J, Sundaram H (2018) The size conundrum: why online knowledge markets can fail at scale. In: Proceedings of the 2018 world wide web conference, pp 65–75. https://doi.org/10.1145/3178876.3186037
12. Santos T, Walk S, Kern R, Strohmaier M, Helic D (2019) Self-and cross-excitation in stack exchange question & answer communities. In: The world wide web conference, pp 1634–1645. https://doi.org/10.1145/3308558.3313440
13. Oliver PE, Marwell G (2001) Whatever happened to critical mass theory? A retrospective and assessment. Sociol Theory 19(3):292–311. https://doi.org/10.1111/0735-2751.00142
14. Smiljanić J, Mitrović Dankulov M (2017) Associative nature of event participation dynamics: a network theory approach. PLoS ONE 12(2):0171565. https://doi.org/10.1371/journal.pone.0171565
15. Török J, Kertész J (2017) Cascading collapse of online social networks. Sci Rep 7(1):16743. https://doi.org/10.1038/s41598-017-17135-1
16. Lőrincz L, Koltai J, Győr AF, Takács K (2019) Collapse of an online social network: burning social capital to create it? Soc Netw 57:43–53. https://doi.org/10.1016/j.socnet.2018.11.004
17. Wasko MM, Faraj S (2005) Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. MIS Q 29(1):35–57. https://doi.org/10.2307/25148667
18. Hung S-Y, Durcikova A, Lai H-M, Lin W-M (2011) The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior. Int J Hum-Comput Stud 69(6):415–427. https://doi.org/10.1016/j.ijhcs.2011.02.004
19. Rode H (2016) To share or not to share: the effects of extrinsic and intrinsic motivations on knowledge-sharing in enterprise social media platforms. J Inf Technol 31(2):152–165. https://doi.org/10.1057/jit.2016.8

20. Kairam SR, Wang DJ, Leskovec J (2012) The life and death of online groups: predicting group growth and longevity. In: Proceedings of the fifth ACM international conference on web search and data mining, pp 673–682. https://doi.org/10.1145/2124295.2124374

21. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: structure and dynamics. Phys Rep 424(4–5):175–308. https://doi.org/10.1016/j.physrep.2005.10.009

22. Gallagher RJ, Young J-G, Welles BF (2021) A clarified typology of core-periphery structure in networks. Sci Adv 7(12):9800. https://doi.org/10.1126/sciadv.abc9800

23. Melnikov A, Lee J, Rivera V, Mazzara M, Longo L (2018) Towards dynamic interaction-based reputation models. In: 2018 IEEE 32nd international conference on Advanced Information Networking and Applications (AINA), pp 422–428. https://doi.org/10.1109/AINA.2018.00070

24. Wei X, Chen W, Zhu K (2015) Motivating user contributions in online knowledge communities: virtual rewards and reputation. In: 2015 48th Hawaii international conference on system sciences. IEEE, pp 3760–3769. https://doi.org/10.1109/HICSS.2015.452

25. Yanovsky S, Hoernle N, Lev O, Gal K (2019) One size does not fit all: badge behavior in q&a sites. In: Proceedings of the 27th ACM conference on user modeling, adaptation and personalization, pp 113–120. https://doi.org/10.1145/3320435.3320438

26. Santos T, Burghardt K, Lerman K, Helic D (2020) Can badges Foster a more welcoming culture on q&a boards? In: Proceedings of the international AAAI conference on web and social media, vol 14, pp 969–973

27. Bornfeld B, Rafaeli S (2019) When interaction is valuable: feedback, churn and survival on community question and answer sites: the case of stack exchange. In: Proceedings of the 52nd Hawaii international conference on system sciences

28. Kang M (2021) Motivational affordances and survival of new askers on social q&a sites: the case of stack exchange network. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/asi.24548

29. Ahmed S, Yang S, Johri A (2015) Does online q&a activity vary based on topic: a comparison of technical and non-technical stack exchange forums. In: Proceedings of the second (2015) ACM conference on learning@ scale, pp 393–398. https://doi.org/10.1145/2724660.2728701

30. Chen G, Mok L (2021) Characterizing growth and decline in online ux communities. In: Extended abstracts of the 2021 CHI conference on human factors in computing systems, pp 1–7. https://doi.org/10.1145/3411763.3451646

31. Posnett D, Warburg E, Devanbu P, Filkov V (2012) Mining stack exchange: expertise is evident from initial contributions. In: 2012 international conference on social informatics. IEEE, pp 199–204. https://doi.org/10.1109/SocialInformatics.2012.67

32. Pal A, Chang S, Konstan JA (2012) Evolution of experts in question answering communities. In: Sixth international AAAI conference on weblogs and social media

33. Oliveira N, Muller M, Andrade N, Reinecke K (2018) The exchange in stackexchange: Divergences between stack overflow and its culturally diverse participants. Proc ACM Hum-Comput Interact 2(CSCW):1–22. https://doi.org/10.1145/3274399

34. Dover Y, Kelman G (2018) Emergence of online communities: empirical evidence and theory. PLoS ONE 13(11):0205167. https://doi.org/10.1371/journal.pone.0205167

35. Han X, Cao S, Shen Z, Zhang B, Wang W-X, Cressman R, Stanley HE (2017) Emergence of communities and diversity in social networks. Proc Natl Acad Sci 114(11):2887–2891. https://doi.org/10.1073/pnas.1608164114

36. Kleineberg K-K, Boguñá M (2015) Digital ecology: coexistence and domination among interacting networks. Sci Rep 5(1):1–11. https://doi.org/10.1038/srep10268

37. Oliver P, Marwell G, Teixeira R (1985) A theory of the critical mass. I. Interdependence, group heterogeneity, and the production of collective action. Am J Sociol 91(3):522–556. https://doi.org/10.1086/228313

38. Dunning D, Anderson JE, Schlösser T, Ehlebracht D, Fetchenhauer D (2014) Trust at zero acquaintance: more a matter of respect than expectation of reward, vol 107 pp 122–141. https://doi.org/10.1037/a0036673

39. Dunning D, Fetchenhauer D, Schlösser T (2019) Why people trust: solved puzzles and open mysteries. Curr Dir Psychol Sci 28(4):366–371. https://doi.org/10.1177/0963721419838255

40. Schilke O, Reimann M, Cook KS (2021) Trust in Social Relations. Annu Rev Sociol 47(1):239–259. https://doi.org/10.1146/annurev-soc-082120-082850

41. McEvily B, Zaheer A, Soda G (2021) Network trust. In: Gillespie N, Fulmer A, Lewicki R (eds) Understanding trust in organizations. Taylor & Francis. https://doi.org/10.4324/9780429449185

42. Aberer K, Despotovic Z (2001) Managing trust in a peer-2-peer information system. In: CIKM'01. Association for Computing Machinery, New York, pp 310–317. https://doi.org/10.1145/502585.502638

43. Duma C, Shahmehri N, Caronni G (2005) Dynamic trust metrics for peer-to-peer systems. In: 16th international workshop on database and expert systems applications (DEXA'05). IEEE, pp 776–781. https://doi.org/10.1109/DEXA.2005.80

44. Tschannen-Moran M, Hoy W (2000) A multidisciplinary analysis of the nature, meaning, and measurement of trust. In: Review of educational research, vol 70. American Educational Research Association, pp 547–593. https://doi.org/10.3102/00346543070004547

45. Dover Y, Goldenberg J, Shapira D (2020) Sustainable online communities exhibit distinct hierarchical structures across scales of size. Proc R Soc A 476(2239):20190730. https://doi.org/10.1098/rspa.2019.0730

46. Orsini C, Dankulov MM, Colomer-de-Simón P, Jamakovic A, Mahadevan P, Vahdat A, Bassler KE, Toroczkai Z, Boguná M, Caldarelli G et al (2015) Quantifying randomness in real networks. Nat Commun 6(1):8627. https://doi.org/10.1038/ncomms9627

47. Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, pp 44–54. https://doi.org/10.1145/1150402.1150412

48. Centola D, Eguíluz VM, Macy MW (2007) Cascade dynamics of complex propagation. Phys A, Stat Mech Appl 374(1):449–456. https://doi.org/10.1016/j.physa.2006.06.018

49. Bollobás B, Riordan OM (2003) Mathematical results on scale-free random graphs. In: Handbook of graphs and networks: from the genome to the Internet, pp 1–34

50. Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174. https://doi.org/10.1016/j.physrep.2009.11.002
51. Saramäki J, Moro E (2015) From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. Eur Phys J B 88(6):1–10. https://doi.org/10.1140/epjb/e2015-60106-6
52. Krings G, Karsai M, Bernhardsson S, Blondel VD, Saramäki J (2012) Effects of time window size and placement on the structure of an aggregated communication network. EPJ Data Sci 1(1):1. https://doi.org/10.1140/epjds4
53. Barrat A, Gelardi V, Le Bail D, Claidiere N (2021) From temporal network data to the dynamics of social relationships. Proc R Soc Lond B, Biol Sci 288:20211164. https://doi.org/10.1098/rspb.2021.1164
54. Arnold NA, Steer B, Hafnaoui I, Parada GHA, Mondragon RJ, Cuadrado F, Clegg RG (2021) Moving with the times: investigating the alt-right network gab with temporal interaction graphs. Proc ACM Hum-Comput Interact 5(CSCW2) 447. https://doi.org/10.1145/3479591
55. Yashkina E, Pinigin A, Lee J, Mazzara M, Adekotujo AS, Zubair A, Longo L (2019) Expressing trust with temporal frequency of user interaction in online communities. In: Advanced information networking and applications. Springer, Cham. https://doi.org/10.1007/978-3-030-15032-7_95