A Method for Identifying Bridges in Online Social Networks

Andrey N. Rabchevskiy[1,2](✉), Victor S. Zayakin[1,3], and Evgeny A. Rabchevskiy[1]

[1] JSC “SEUSLAB”, Perm, Russia
andrey@ranat.ru
[2] Perm State University, Perm, Russia
[3] National Research University “Higher School of Economics”, Perm, Russia

Abstract. The current level of development of online social networks has transformed social media from a way of communication between people into a tool for influencing people’s behaviour in their daily lives. This influence is often aimed at inciting protest movements in society and mobilising citizens for protest actions, and has a targeted impact on social network users. The sponsors and main actors of disruptive influences are often forces located in other countries. In the context of counteraction to targeted destructive influences, the task of identifying the network structure of destructive influence is very relevant. One element of this structure is the users connecting individual communities to the core of the protest network. These users are the bridges between the clusters and the core network. Their main task is to contribute to the rapid growth of the protest audience. Identifying the most influential bridges and blocking them could decrease the protest potential or make the protest actions ineffective. In this paper, we propose a methodology for identifying bridge users based on the original centrality measure of weighted contribution. Moreover, a method for identifying the most influential bridges is proposed. Unlike most probabilistic methods, weighted contribution centrality allows for clear determination of whether a user is a bridge or not. A description of the measure, a mathematical model and an algorithm for calculating it are presented.

Keywords: Online social networks · Social network analysis · Structure of protest network · Core of protest network · Clusters · Community · Bridges · Weighted contribution centrality

1 Introduction

Today’s social networks are no longer just a means of communication between people and have evolved into an effective tool for targeting users. The aim of influence can be to engage users in specific thematic communities or to disseminate information that can influence people’s behaviour in everyday life.

E. Burnaev et al. (Eds.): AIST 2021, CCIS 1573, pp. 166–175, 2022.
https://doi.org/10.1007/978-3-031-15168-2_14

Examples of these influences are the political events of the Arab Spring in 2010–2011, the #Occupay movement in the US in 2011, the protests in Turkey, Brazil and Hong Kong (2013–2014), the recent presidential elections in Belarus (2020) and the political actions around the arrest of Navalny and "Putin's palace" in 2021, where social media were used to coordinate people into actual political actions.

The study of the mechanisms and degree of influence of social networks on people's behaviour has generated a great deal of scientific interest. According to [1–3], all protest movements are inextricably linked to the creation of autonomous communication networks supported by the Internet. The significant impact of social networks on the level of people's mobilization for action has been described in [4–6]. When studying social networks in the context of protest sentiments, one often observes their pronounced cluster structure. Figure 1 shows examples of graphs of such networks, where the vertices of the graph are users and the edges are connections between them. The colours indicates the level of publication activity of users in the social network, i.e. the number of any type of material on the target topic published by the user. Red indicates the maximum level of publication activity and grey indicates no activity.

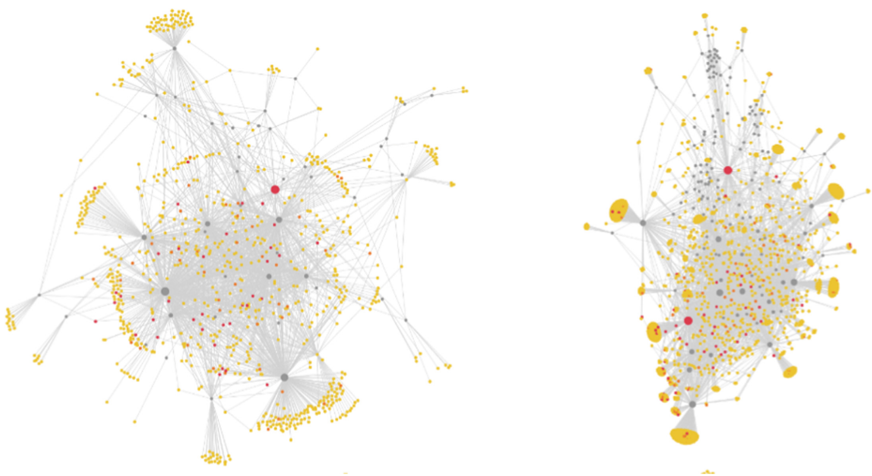

Fig. 1. Examples of the cluster structure of social network graphs.

As can be seen from Fig. 1, most of the graphs have a pronounced cluster structure in the form of a core with many cross-links between users and isolated clusters that are connected to the core through a single user acting as a "bridge" between the cluster and the core.

Since the publication activity of users in clusters is similar to core users, it is logical to assume that the sum of the activity levels of each node in a cluster can be higher than that of any node in the core, and the node connecting the cluster to the core will contribute more to the overall network activity level than any node in the social network core. An analysis of the profiles of users acting as bridges in the protest activity theme showed the following results:

– users in clusters often only partially share the views of core users on the objects of discussion;
– users in clusters are often united by the same topic;
– the preferences, interests, and political views (for protest networks) of users in different clusters may differ (they may belong to different political parties or movements), but these users share opposition to the current authorities;
– as a rule, users connecting the cluster to the core act as community moderators;
– such users have connections with each other and form a substructure in the graph of social relations, due to which they can coordinate their actions, involving in the social phenomenon under study different categories of users, possibly disagreeing with the common point of view of the network core on some issues.

Thus, identifying users who act as social media bridges is crucial to counteracting protest movements and managing the parameters of the spread of viral and destructive information on social media.

Using the software "SEUS search engine" [7], actively used by law enforcement agencies of the Russian Federation [8], we searched for publications in the social network VKontakte related to the organization of protest events in January–August 2019 in Moscow. For each user the level of publication activity was calculated, which took into account the number of posts, reposts, comments, likes, etc. As a result of ranking by the level of publication activity, a ranking of user activity was compiled. For each user the graphs of the social connections of the users' friends and friends of their friends were constructed. The following conditions were taken into account:

– a user is included in the graph if he is a friend of a member of the activity rating or is a friend of any of his friends (the maximum distance to the target user in the graph is two);
– a user whose activity level is zero is included in the graph only if he is a friend of at least two users from the activity rating.

A social network node that satisfies the following requirements was considered a bridge:

– a node that connects the cluster to the core of the network;
– cluster nodes are only connected to the bridge and are not connected to each other;
– bridge is connected to cluster nodes and core nodes.

Figure 2 shows a fragment of a typical node acting as a bridge.

In graph theory these nodes are usually called articulation node, cut-node or broker, but we will use "bridge" for ease of reference. Thus, the challenge was to select or develop a methodology that unambiguously identifies bridges in cluster networks and also determines the extent to which bridges influence the overall level of network activity.

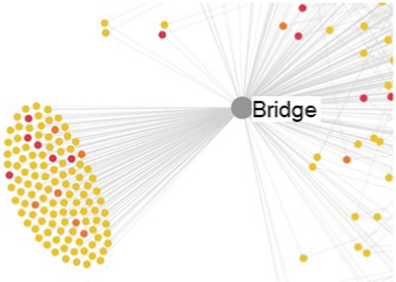

Fig. 2. Fragment of the user-bridge graph.

2 Materials

Various centrality metrics are used to identify the characteristics of nodes in networks, as described in the review paper [9]. The best known centrality measure that characterizes the communicative ability of nodes is the centrality on betweenness, first introduced independently [10] and [11] and finally formulated in [12]. Betweenness corresponds to the sum of all the shortest paths that pass through a given node in the graph. Since each node in the networks we studied had a certain level of activity, it was necessary to consider the weight of each node in the network. To calculate intervening centrality for weighted networks, the techniques proposed in [13–17] could be used. However, betweenness centrality, with or without weight, can reveal the level of communication capability of a node in the network, but cannot accurately determine whether a given node is truly a bridge, since nodes with a high betweenness centrality value can be located both at the core of the network and at the periphery of the network, being bridges.

Influential nodes, according to [18], always act as a "bridge" between communities and exist within an overlapping community. The authors suggest using the local centrality method to identify such influential nodes, which assumes that the more communities a node belongs to, the more influence it has. In [19, 20], "transmission centrality" and "modular centrality" measures are proposed to define bridges, but transmission centrality can be high in both core nodes and bridges, so its meaning is not very different from intermediate centrality, and in modular centrality, nodes connecting communities are the bridge, whereas we investigated nodes between communities and the network core, meaning that the concept of bridge had a different meaning in this context.

A method that successfully identifies bridges is presented in [21], in which the authors introduced the concept of Bridging Centrality. This measure identifies bridges more accurately, but it works only in sparse networks with a large number of bends, because it is based on the idea that to identify bridges it is necessary to discard the value of links with nodes that are in close proximity to a node, that is, links of the first knee of the graph. Since, in our case cluster users are connected only with a bridge, they cannot be taken into account in the calculation of this measure, which did not suit us. The closest measure for our problem is "Contribution centrality" proposed by [22], the essence of which is that the centrality of a node is proportional to the sum of centrality of nodes in its neighborhood, weighted by their contributions. The contribution centrality is indeed the most applicable for our problem, since it can determine the contribution of each bridge for the kernel users, but it does not guarantee an unambiguous definition of the bridge, which in our case was a necessary condition.

As we can see, all the measures presented above could, to a greater or lesser extent, determine the level of communication capability of a node, but cannot exactly determine whether a given node is a bridge as we understand it.

3 Method

We will say that all users with publication activity on the topic of a given social phenomenon and their social connections constitute the "temporary social network" generated by this social phenomenon, and the sum of the activity levels of all users constitutes the total activity level of the temporary social network.

Since the number of users in different clusters and their level of activity are different, bridges can have different levels of influence. Let the degree of influence of the bridge on the overall level of publication activity of the temporary social network be defined as the total level of activity of the cluster that is connected to the core through the bridge. According to the above definition of bridge, cluster nodes should only be connected to the bridge and should not be connected to each other. Consider the graph shown in Fig. 3 and calculate which of the nodes in the graph is a bridge in the context of the proposed definition and calculate its cluster weight.

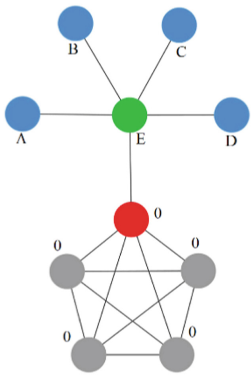

Fig. 3. Node network diagram including weights.

For the red node, links with the green and grey nodes will not be taken into account as they have links with other nodes in the network, so the value of the red node's total rating, as well as for the grey nodes, will be 0.

For blue nodes which only have links with the green node, the value will also be 0 as the green node also has other links,

For the green node the link with the red node will also give 0, and the links with the blue nodes will give the value of the weights of those nodes,

The weight of the green node will be equal to the sum of the weights of the blue nodes $E = A + B + C + D$.

As we can see, we obtained a single non-zero value for the green node in the whole network, which exactly determines the presence of the sought bridge and its contribution to the overall level of activity of the temporal social network, equal to the value of E.

Given that the weight of each node in the graph plays a significant role in the proposed method, the term "Weighted Contribution Centrality" was proposed to determine the degree of influence of the bridge on the overall level of activity.

We will say that weighted contribution centrality is the contribution of a social network node to the total level of publication activity, equal to the sum of the activity of each cluster node connected by the node to the core network, divided by the total activity level of the network. In other words, bridge weighted contribution centrality is the weight of the cluster connected by the bridge to the core, divided by the total weight of the network.

The weighted contribution centrality value = 0 if the node is not a bridge, and > 0 if the node is a bridge. The most influential node in the bridge role has the highest weighted contribution centrality value for this network. Let us introduce notations to formally describe the proposed methodology.

Let $G = (U, F)$ be a graph consisting of a set U of users and a set F of disordered pairs of different elements of the set U, reflecting friendly relations between users (graph edges).

If users u and υ are friends, i.e. form a relation $f \in F$, we write $f = (u\upsilon) = (\upsilon u)$. Denote the set of friends of user $u \in U$ by $F(u) = \{\upsilon \in U\colon (u\upsilon) \in F\}$. Then the degree of a node, i.e. the number of friends of user $u \in U$, is naturally denoted by $|F(u)|$. The set of users associated only with a given user u is called the neighbours of user $u \in U$ and denoted by $S(u)$. Then:

$$S(u) = \{\upsilon \in U : (u\upsilon) \in F, |F(\upsilon)| = 1\} = \{\upsilon \in U : |F(\upsilon)| = 1\} \cap F(\upsilon) \quad (1)$$

If user activity level (i.e. the number of publications on the target topic) $u \in U$ is denoted by $r(u)$, then the total activity level of some subset of users $V \subset U$ will be calculated by the formula:

$$R(V) = \sum\nolimits_{u \in V} r(u) \quad (2)$$

Using formula (2), we get a formula for_calculating the weight of a bridge-connected cluster of an arbitrary graph user:

$$C_{WC}(u) = R(S(u)) = \sum\nolimits_{w \in S(u)} r(w) \quad (3)$$

Weighted contribution centrality is defined as the ratio of the weight of the cluster connected by the bridge to the network core to the total activity level of all clusters in the network. The total activity level of all clusters in the network can be calculated as

$$R = \sum\nolimits_{u \in V} W(u) \quad (4)$$

Thus, weighted contribution centrality can be expressed as

$$C_{WC}(u) = \frac{W(u)}{R} \quad (5)$$

The code for the Python3 function used to calculate bridges as part of the Python program [23] is shown below:

```
def calculate_weighted_contribution_centrality(graph, rating):
    centrality = {}
    R = 0 # R = 0 # accumulated weight value of all clusters
    for user, friends in graph.items():
        c = 0 # accumulative value of cluster weight to be connected by user
        # accumulate rating by user's friends
        for friend in friends:
                                        # if a user's friend is linked in the column only, add their
                                     rating
            if len(graph[friend]) == 1:
                c = c + rating[friend]
        centrality[user] = c
        R = R + c
    for user in centrality.keys():
        centrality[user] = centrality[user] / R
    return centrality
```

3.1 Evaluating the Effectiveness of the Bridge Detection Method

To determine the level of influence of bridges from 10 random graphs, the 10 most influential bridges and their associated vertices were removed, as well as those vertices that were isolated after the bridges were removed. Table 1 shows how much the weight of the graphs as a whole and the total weight of the vertices included in the clusters decreased.

As Table 1 indicates, when the 10 most influential bridges are removed from the graphs, the total weight of the graph or the total level of user activity in the graph decreases by an average of 57.8%, indicating a high level of influence of the bridges. At the same time, the total cluster weight decreases by 80.9%, which corresponds to the role played by the 10 most influential bridges in network expansion. From this we can conclude that the network nodes we identified as bridges do contribute significantly to the overall level of network activity. A comparison of the results obtained using centrality by intermediacy and centrality by contribution is presented in (Table 2).

As shown in Table 2, when the 10 most influential bridges along with all their nodes are removed from the graphs, the total weight of the graph decreases approximately equally. This suggests that all three measures are equally effective in revealing the communication abilities of the influential nodes in the network. At the same time, the change in cluster weight is noticeably larger when using weighted contribution centrality.

Table 1. Change in graph weight as a result of removing 10 bridges with their vertices from the graph.

Graph number	Changing graph weight	Changing cluster weights
1	−32,7%	−67,5%
2	−53,0%	−88,6%
3	−44,1%	−65,9%
4	−92,0%	−94,5%
5	−97,3%	−99,7%
6	−72,1%	−92,6%
7	−21,9%	−61,0%
8	−27,2%	−68,3%
9	−48,5%	−72,3%
10	−83,4%	−98,9%
	Average value	*Average value*
	−57,2%	**−80,9%**

Table 2. Change in various centrality measures as a result of removing 10 bridges with their vertices from the graph.

Measure	Changing graph weight	Changing cluster weights
Betweenness centrality	−54,4%	−76,0%
Contribution centrality	−52,7%	−66,4%
Weighted contribution centrality	−57,2%	−80,9%

This is because betweenness centrality and contribution centrality identify the most communicative nodes in the network, including bridges, as opposed to weighted contribution centrality, which only identifies bridges. And since the removal of bridges gives the largest contribution to the reduction in the overall level of network activity, the impact of bridges is greater than that of any other nodes in the network.

Thus, it can be argued that centrality on weighted contribution solves the bridging problem most effectively compared to the other metrics presented.

4 Conclusion

A feature of the weighted contribution centrality measure is that it unambiguously determines whether a node is a bridge in the network configurations described earlier.

Bridges contribute to expanding the size of the network, increasing the number of users involved in the social phenomenon and increasing the overall level of activity of the social network. Blocking the most influential bridges can significantly change the characteristics of the entire network and reduce the overall level of social network activity in a given social phenomenon. Therefore, targeting the most influential bridges is an effective way to reduce social network activity.

The level of informational influence is assessed by ranking the bridges in order of centrality by weighted contribution.

This method of identifying bridges and assessing their informational impact was used as part of an analytical study "political protest propaganda structures in Russia and Belarus", conducted by the SEUSLAB analytical centre LLC. The research included an assessment of the operational significance of the findings and it was presented on the site of the CIS Antiterrorist Centre and the CIS Research Institute for Security Problems in March 2021. Based on the results of the piloting, the Scientific Advisory Board of the CIS Antiterrorist Center drafted an expert opinion on the feasibility of using this method in the information and analytical systems used in the operational and service activities of the Russian Interior Ministry.

References

1. Castells, M.: Networks of Outrage and Hope Social Movements in the Internet Age. Polity, Cambridge (2012)
2. Gerbaudo, P.: Tweets and the Streets. Social Media and Contemporary Activism. Pluto Books, London (2012)
3. Faris, D.M.: Dissent and Revolution in a Digital Age. I.B.Tauris, London (2013). https://doi.org/10.5040/9780755607839
4. Tindall, D.B.: From metaphors to mechanisms: critical issues in networks and social movements research. Soc. Netw. **29**, 160–168 (2007). https://doi.org/10.1016/j.socnet.2006.07.001
5. Bennett, W.L., Segerberg, A.: The logic of connective action. Inf. Commun. Soc. **15**, 739–768 (2012). https://doi.org/10.1080/1369118X.2012.670661
6. Juris, J.S.: Reflections on #Occupy everywhere: social media, public space, and emerging logics of aggregation. Am. Ethnol. **39**, 259–279 (2012). https://doi.org/10.1111/j.1548-1425.2012.01362.x
7. https://www.seuslab.ru/seus
8. https://meduza.io/feature/2018/10/16/politsiya-po-vsey-rossii-pokupaet-sistemy-monitoringa-sotssetey-oni-pomogayut-iskat-ekstremizm-ne-vyhodya-iz-rabochego-kabineta
9. Lü, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C., Zhou, T.: Vital nodes identification in complex networks. Phys. Rep. **650**, 1–63 (2016). https://doi.org/10.1016/j.physrep.2016.06.007
10. Anthonisse, J.M.: The Rush in a Graph. Mathematisch Centrum (mimeo) (1971)
11. Freeman, L.C.: A set of measures of centrality based on betweenness. Sociometry **40**, 35 (1977). https://doi.org/10.2307/3033543
12. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**, 215–239 (1978). https://doi.org/10.1016/0378-8733(78)90021-7
13. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: generalizing degree and shortest paths. Soc. Netw. **32**, 245–251 (2010). https://doi.org/10.1016/j.socnet.2010.03.006
14. Kuznetsov, E.N.: Analysis of the structure of network interactions: context-dependent measures of centrality. Management of large systems, pp. 57–82. IPU RAS, Moscow (2019)

15. Wang, H., Hernandez, J.M., van Mieghem, P.: Betweenness centrality in a weighted network. Phys. Rev. E **77**, 046105 (2008). https://doi.org/10.1103/PhysRevE.77.046105
16. van Mieghem, P., van Langen, S.: Influence of the link weight structure on the shortest path. Phys. Rev. E. **71**, 056113 (2005). https://doi.org/10.1103/PhysRevE.71.056113
17. Levandowsky, M., Winter, D.: Distance between sets. Nature **234**, 34–35 (1971). https://doi.org/10.1038/234034a0
18. Wei, H., et al.: Identifying influential nodes based on network representation learning in complex networks. PLOS ONE **13**, e0200091 (2018). https://doi.org/10.1371/journal.pone.0200091
19. Zhang, Q., Karsai, M., Vespignani, A.: Link transmission centrality in large-scale social networks. EPJ Data Sci. **7**(1), 1–16 (2018). https://doi.org/10.1140/epjds/s13688-018-0162-8
20. Ghalmane, Z., El Hassouni, M., Cherifi, C., Cherifi, H.: Centrality in modular networks. EPJ Data Sci. **8**(1), 1–27 (2019). https://doi.org/10.1140/epjds/s13688-019-0195-7
21. Jensen, P., et al.: Detecting global bridges in networks. IMA J. Complex Netw. **4**, 319–329 (2015)
22. Alvarez-Socorro, A.J., Herrera-Almarza, G.C., González-Díaz, L.A.: Eigencentrality based on dissimilarity measures reveals central nodes in complex networks. Sci. Rep. **5**, 17095 (2015). https://doi.org/10.1038/srep17095
23. Rabchevskiy, A.N., Zayakin, V.S.: The program for calculating bridges in cluster networks. Certificate of state registration of computer programs No. 2021616086 of 16 April 2021 (2021)