Finding Outcasts in Social Networks

Developed by Matias Ijäs, Finnish Defence Research Agency

Contact Author Email: matias.ijas@mil.fi, matias.ijas2@gmail.com

Introduction

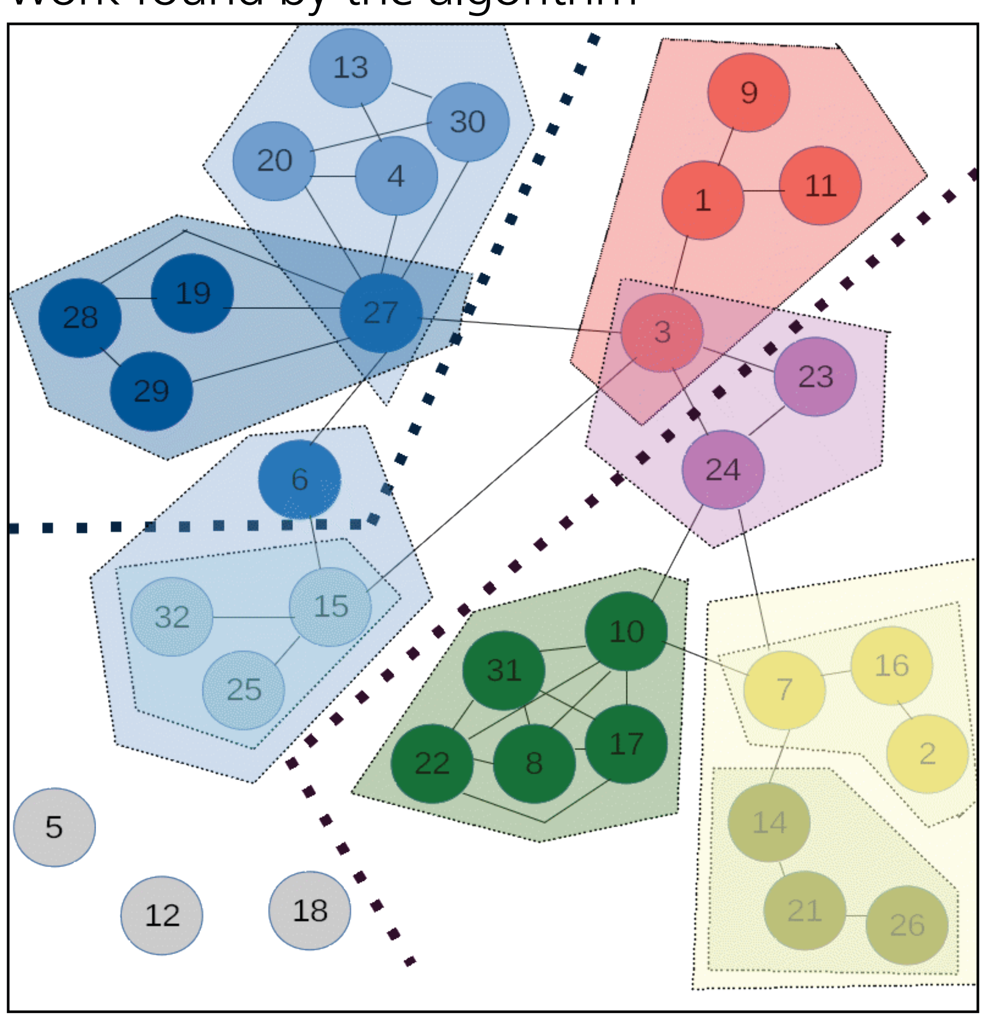
Social network analysis has brought great advances to our understanding of complex systems. In this poster the algorithm of finding outcasts is represented. The algorithm consists of 4 different approaches and provides answers for multiple problems.

Fig. 1. Social exclusion is a problem in all age groups

Social exclusion is often provoked by government and people's prejudices. Most popular reasons behind social exclusion are cultural differences, gender, ethnicity and age. Very often tragic event triggers exclusion from the social communities.

The model and algorithms are created to find outcasts in social networks, which represent different enviroments, like school, workplace, social media or other communities in real life.The goal is to create a working model, which can find outcasts from the group with accuracy of 80% or greater

Fig. 2. The groups of the Dutch student network found by the algorithm



Model & Algorithm

The model analyses the risk of being isolated by computing spreading probabilities and node's popularity in groups. The less influence the node has and fewer groups node belongs to, the higher percentage of outcast risk the node has.

Fig. 3. The equations of Closeness centrality (C_c), Betweenness Centrality (C_b) and Outcast index (O_i)

$$C_c(i) = \frac{\sum_{j \neq i} g_{i,j}^{-1}}{N-1}$$

$$C_b(v) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$$O_i(n) = \frac{1 + \frac{\sqrt{S_{gmax}}P_{gmax} - \sqrt{S_n}P_n}{\sqrt{S_{gmax}}P_{gmax} + \sqrt{S_n}P_n} + \frac{k}{N}\sum_{t=1}^{N} P_{s,t}}{2+k}$$

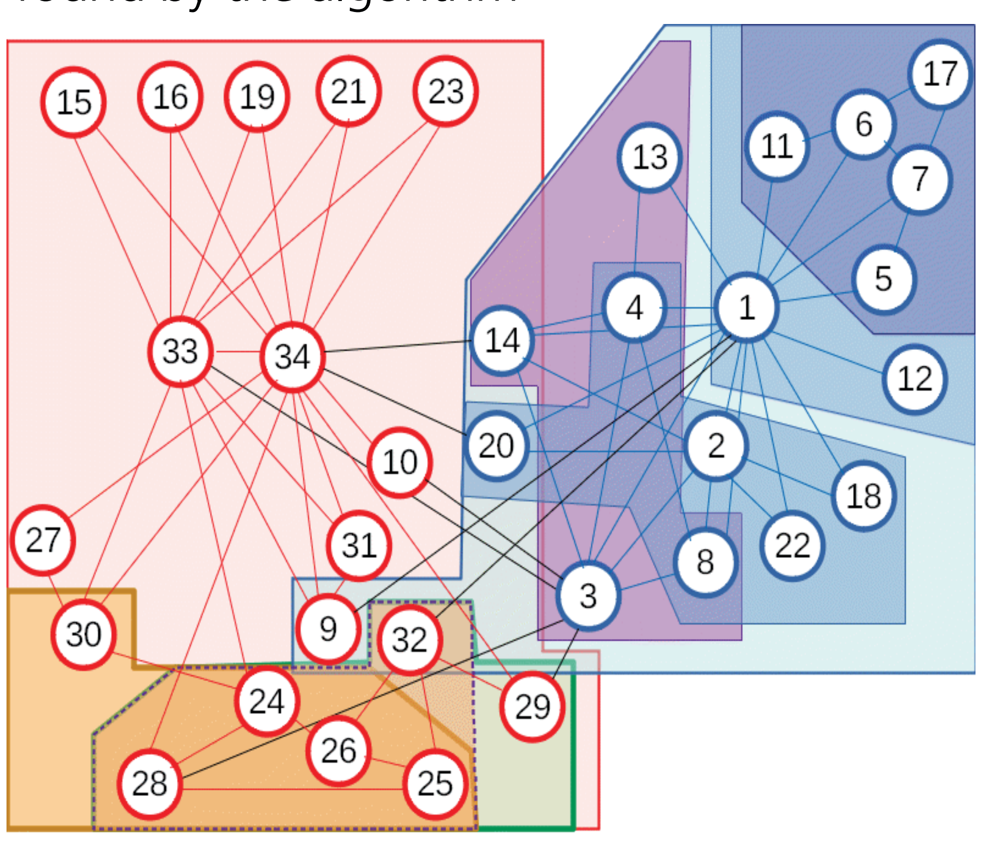
The model is based on 4 different approaches in Network Science.

- Closeness centrality
- Betweenness centrality
- Node degree
- Community detection

Those are used in our combined algorithm, which has 4 main phases.

- Converting the network data
- Computing centrality measures
- Finding communities
- Analyzing computed values

Fig. 4. The groups of Zachary's karate club found by the algorithm

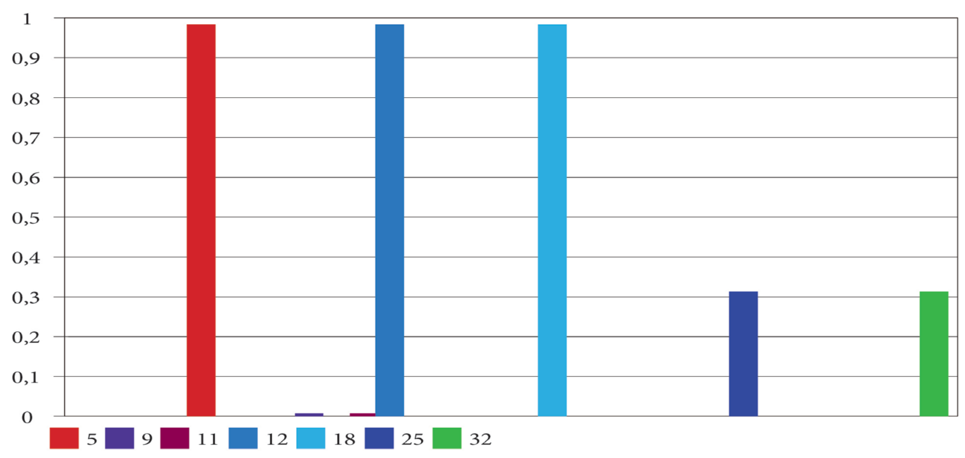


Test data & Results

Three different public datasets were used; Dutch network of 32 students, Zachary's karate club with 36 nodes and Facebook with 4039 nodes.

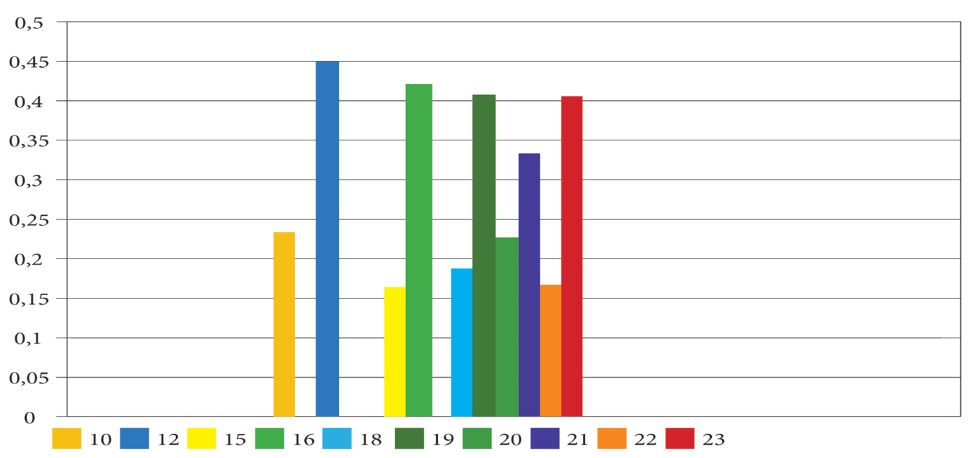
In the Dutch student network, 3 students (5, 12, and 18) had no links to any other node, so obviously they would be isolated. The algorithm gave them 98% risk of being isolated. Other nodes that have risk to be isolated are 25 (31%), 32 (31 %), 9 (<1%) and 11 (<1%).

Fig. 5. Risk of being isolated in Dutch student network



In the Zachary's karate club, results show that the nodes 12 (45%), 16 (42%), 19 (40%) and 23 (40%) have highest risks of being isolated. There were 6 other nodes (10, 15, 18, 20, 21, 22), which had risk between 14 and 33 percent of being isolated.

Fig. 6. Risk of being isolated in Zachary's karate club network



In the Facebook network, which is a larger network, the nodes 911, 918, 1096, 1119, 1145, 1206, 1386, 1466, 1560, 1581, 1834 and 3071 had highest risk of being isolated with the probability of 57,8%. There were 216 nodes, which had 10% or greater probability of being isolated and 68 of those had probability over 50%.

Conclusion

In simple networks, such as Zachary's karate club and Dutch students, the results are as expected and similar to when calculating by node degrees. Larger networks, such as Facebook are harder to analyze, whether the result is correct or not, because no other research in this narrow area have been presented in the literature.

In the future, this model and algorithms can be improved in the scalability, performance and accuracy. Different approaches can be made and it would be interesting to analyze the results.

Finnish Defence Research Agency

Puolustusvoimat

Puolustusvoimat.fi