Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

Analytics Using STINGER

CHALLENGE

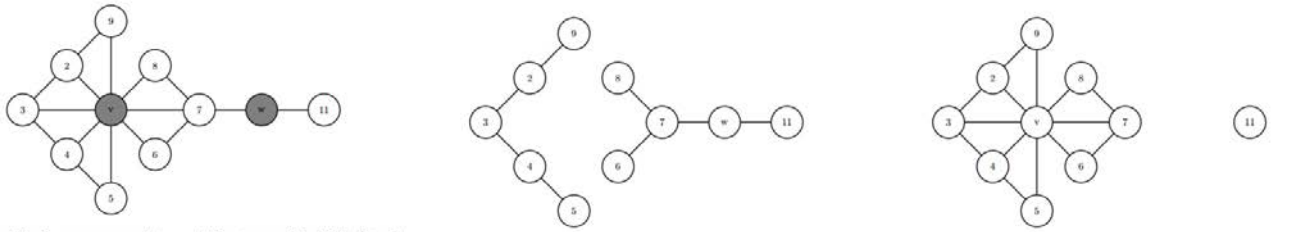
Numerous centrality measures exist and help to quantify the importance of a vertex with respect to the entire graph. These network-wide analytics tend to be both computationally expensive to calculate (limiting scalability) and overlook important players in a local environment. Finding these players requires computationally efficient new metrics that identify both global and local key players.

Developing computationally efficient metrics to monitor important players in near real-time detection for insider-threat and anomaly detection.

CURRENT PRACTICE

Analysts use various centrality measures to quantify the importance of a vertex. Computing these metrics is prohibitive as the size of networks continues to scale and grow beyond billions of members. To handle such large-scale networks, a computing cluster with tens of thousands of processors may be required to analyze a single snapshot of the network. Given the sheer size of the network and the rapid rate at which it changes, by the time the computation has finished analyzing the snapshot of the data, those results may no longer be relevant.

Many widely used analytics focus on finding key players in the entire network, but these analytics tend to overlook local key players because they have different properties from those exhibited by global key players. Yet, local key players can play a significant role. One example of local key players not likely to be found by network-wide analytics is community organizers that hold a community together through their connection and participation in community events. We aim to design algorithms to perform a finer-level analysis and find such players.



(a) Initial community with two highlighted vertices: v and w.

(b) Vertex v is removed from community.

(c) Vertex w is removed from community.

Example of our community-centric analysis. (a) Initially, the graph is decomposed into communities. In this sub-figure, only one community is shown. Using the community-centric analysis approach, each vertex and its respective edges are removed from the initial community. We can then proceed to mark important vertices using our community-centric metrics. We detect that "v" is important (b), while "w" is not (c).

TECHNICAL APPROACH

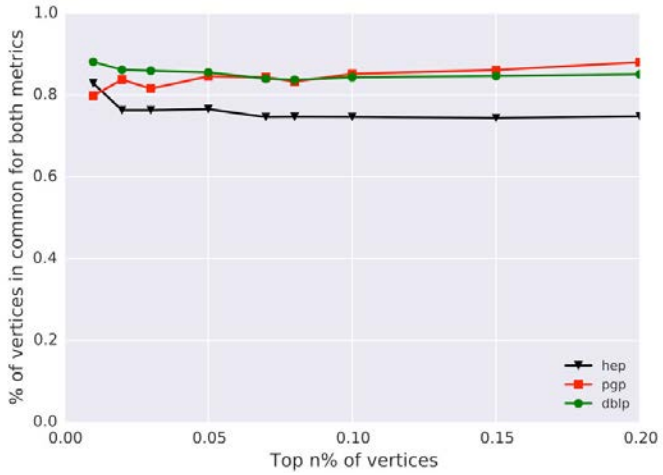
The approach will help pinpoint local key players in a community setting by splitting the original network into communities. Once the local key players are detected, we can create additional capabilities to track these players in a streaming environment where the underlying network is constantly changing.

In addition to finding local key players that are responsible for maintaining the strength of the community, the research team seeks bridge-vertices (bridge players) in the graph that are responsible for maintaining connections with other communities. Using similar techniques for finding local key players, the research team will identify bridge vertex properties to track these vertices in a near-real-time environment.

One major benefit of the proposed community-centric approach is its reduced computational requirements. Specifically, using an analytic at the community level is significantly faster than using the same analytic at the network level. The reduction in the computational requirements is partially due to the reduction in the data set size. Instead of an analytic requiring the whole graph, it now needs only a subgraph, which requires fewer traversals. This approach uses streaming graph analytics in two ways. The first approach uses streaming graph analytics to monitor the change to the underlying network as updates occur. The second uses streaming graph analytics in a controlled fashion to add and remove new relationships between players—allowing us to measure their significance.

Network	Vertices	Edges	Global (sec.)	Local (sec)	Speedup
hep	8k	15k	4.54	0.21	21.62×
PGP	11k	24k	16.38	1.19	14.89×
coAuthors DBLP	540k	15M	94.23	27.6	3.41×

Performance comparison of global key player approach with a local key player approach. Performance is compared as a function of the execution time for each of these approaches. As expected, the local key player approach is faster because of its lower computational requirements.



Global key player versus local key player ranking correlation for betweenness centrality. The x-axis depicts the top percentage of ranked players for both approaches using the betweenness centrality metric. The y-axis depicts the percentage of the vertices marked by both approaches. When asked to give the top 1 percent of the key vertices in the network, both approaches agree on around 85 to 90 percent of the key vertices.

IMPACT

The ability to monitor important players in near real time is useful for a large number of applications, including insider threat and anomaly detection. Currently, this is not feasible because of computational requirements. The described approach will enable analysts to monitor the underlying network with improved accuracy and at a faster rate.

Contacts

Oded Green
Principal Investigator
(404) 385-3386
ogreen@gatech.edu

David A. Bader
(404) 385-4785
bader@cc.gatech.edu

John R. Johnson
Program Director
(509) 375-2651
John.Johnson@pnnl.gov

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Georgia Institute of Technology