Supplementary Material A:

Finding Reach of Dominant Nodes with NetProp Algorithm in Some Benchmark (Undirected) Acquaintance Networks

In this section, we evaluate the ***NetProp*** algorithm on two publicly available, undirected real-world networks selected from the literature on Network Science. These networks were selected because they have been well well-known and studied by many researchers from different perspectives. Here we shall examine the 'flow' of influence from nodes with high dominance centrality towards those nodes that are dominated, and trace the progress of influence spatially through a community of propagators, as discovered by the NetProp algorithm. We shall attempt to show that the story of influence flows that Netprop tells hasn't been previously told. The networks we consider and the corresponding analysis are clubbed together below.

A) The Zachary Karate Club (ZKC) Network

We shall first revisit this popular network science example despite the fact that it has often been thought to leave a question unanswered with its structure. The Zachary Karate Club (ZKC) Network dataset is well known and contains the network of friendships between members of a karate club at a US university, as described by Wayne Zachary in 1977 [1].

As is well known the ZKC data presents an acquaintance or friendship network among 34 individuals who were original members of the club. Since the members later formed two separate groups due to a disagreement among themselves, leaving two dominant (influencing) individuals, both teachers at the club, as the leaders of the two groups, around whom the two communities formed, this data set has often been taken to represent a useful 'before and after' scenario to test and evaluate community detection algorithms, even though the assumption that degree connectivity and/or betweenness centrality alone may have influenced individual choice on whom to support, is thought to be unrealistic. We are of course not interested here with verification of this ground truth by any community detection scheme.

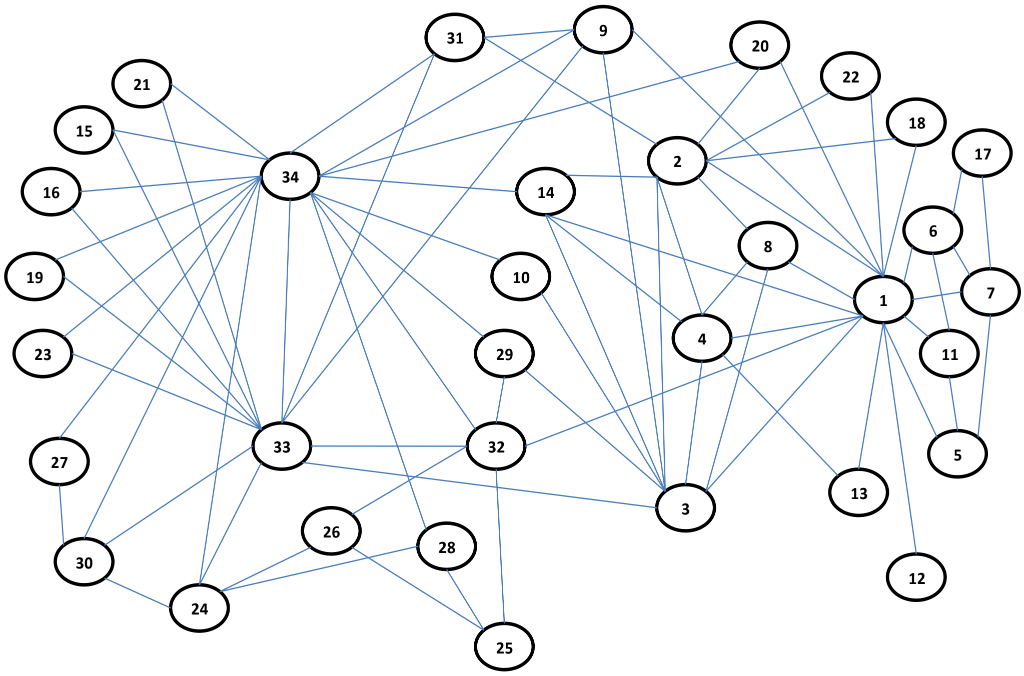
Even so, this data set is useful here to demonstrate the power of DONEX in bringing out how the notion of dominance or influence flows can provide new insights on the propagation of influence and reach of powerful individuals. We shall see later that these notions are easily extended to more complex networks, as well.

Lets start with a visualization of the basic undirected network of the 34 members of the Karate Club shown in Figure A1. The values of DONEX scores together with node degrees for all the 34 nodes are tabulated below the network diagram in Figure 9. Clearly, nodes 34 and 1 are the top two dominant (SDG) nodes (in that order). It may be observed that these individuals dominate over their neighbourhoods, because most such neighbours, although interconnected, have far poorer number of friends themselves. For the purpose of calculation of DONEX, we have assumed that the strengths of relationships, the interaction weights, are all set to 1. Other studies too show that nodes 34 and 1 are the key leaders around whom the members formed the two disjoint communities, DONEX confirms that they are important for their position of dominance.

Using the adjacency data, it is easy to apply traditional community detection algorithms to see how members formed communities around these two individuals on the basis of the edge-betweennes metric. As an example, if we applied the Newman-Girvan algorithm[2], which recursively removes edges that have the highest betweenness until two disconnected communities are formed, we would find the structure shown in Figure 10.

Note that the NG algorithm used here employs (random walk) betweenness measures derived from degree connectivity information in the adjacency matrix for the network. Incidentally, this partition agrees with ground truth. There is, of course, no concept of flows in the NG algorithm. However, if we calculated the dominance flows using DONEX across the edges that disconnect the two communities found with the NG algorithm, we would find a 'boundary flow' of 0.1959, as depicted by the red arrow in the Figure A2, flowing from the community surrounding node 1 towards the community surrounding node 34.

In order to see the impact of derived directions of the edges, and the calculated flows, let us apply NetProp to the data, seeding it first with Node 1, and then separately with Node 34, to determine the minimum flow propagator communities around these individuals. We may then be able to compare minimal flow cuts with minimum betweenness cuts obtained from traditional community detection algorithms.



node	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
deg	16	9	10	6	3	4	4	4	5	2	3	1	2	5	2	2
donex	2.0154	1.2232	1.2802	0.9855	0.9081	0.9928	0.9928	0.8573	0.888	0.8386	0.9081	0.8772	0.8524	0.8848	0.8397	0.8397

node	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
deg	2	2	2	3	2	2	2	5	3	3	2	4	3	4	4	6	12	17
donex	0.8952	0.8443	0.8397	0.8343	0.8397	0.8443	0.8397	0.977	0.8992	0.8973	0.8687	0.9206	0.8391	0.9187	0.8578	1.0598	1.5402	2.1004

Figure A1. Zachary's Karate Club Friendship Network with calculated values of DONEX

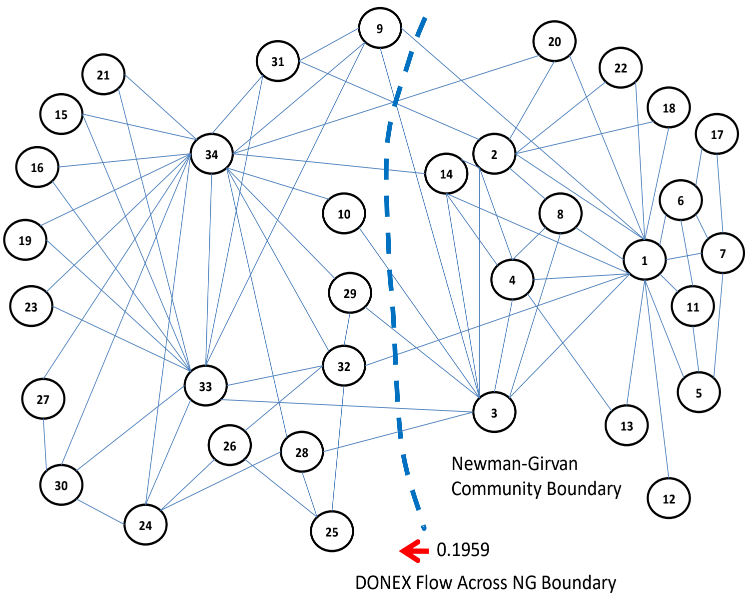


Figure A2. Zachary's Karate Club network partitioned into two communities by the Newman-Girvan Algorithm. The results from the application of NetProp to ZKC with Node 1 as the seed are shown in Figure A3 and A4.

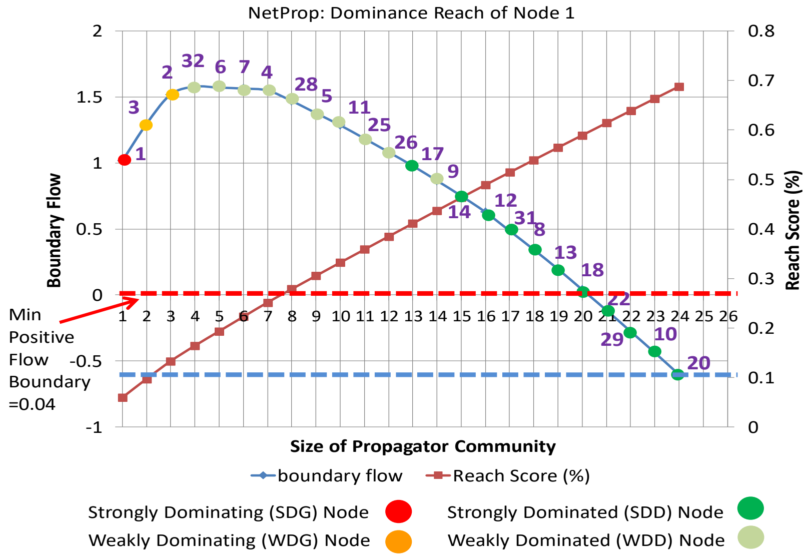


Figure A3. Boundary Flow and Reach Score with enlarging propagator community of Node 1

The plot in Figure A3 shows that boundary flows, beginning from Node 1 (an SDG) increases as expected until nodes 3 and 2 are recruited into the propagator community then stabilizes until the next four WDDs are added. The boundary flows remain positive until node 18 is added, after which it turns negative. At this stage the Reach Score of node 1 is about 60% and the minimum positive flow at the boundary of the propagator community is about 0.04. Recall that RS isn't merely a percentage of the number of nodes of the network, but the percentage of dominance scores over the maximum. Four more nodes, 10,20,22 and 29 which are all SDDs can yet be reached, but they do not add positive values, even though the RS may be improved to about 70%.

This scenario is depicted with the network diagram show in Figure A4. Note the light yellow partition of the network around Node 1 represents the minimum positive flow boundary of the propagator community of Node 1. The pink boundary around this region represents the extent of reach of Node 1, with the addition of four nodes, 10, 20, 22 and 29. Observe that all edges incident on this (pink) partition from the outside are incoming – implying no further dominance flows would be possible. Hence the community representing the partition in blue cannot be reached by Node 1.

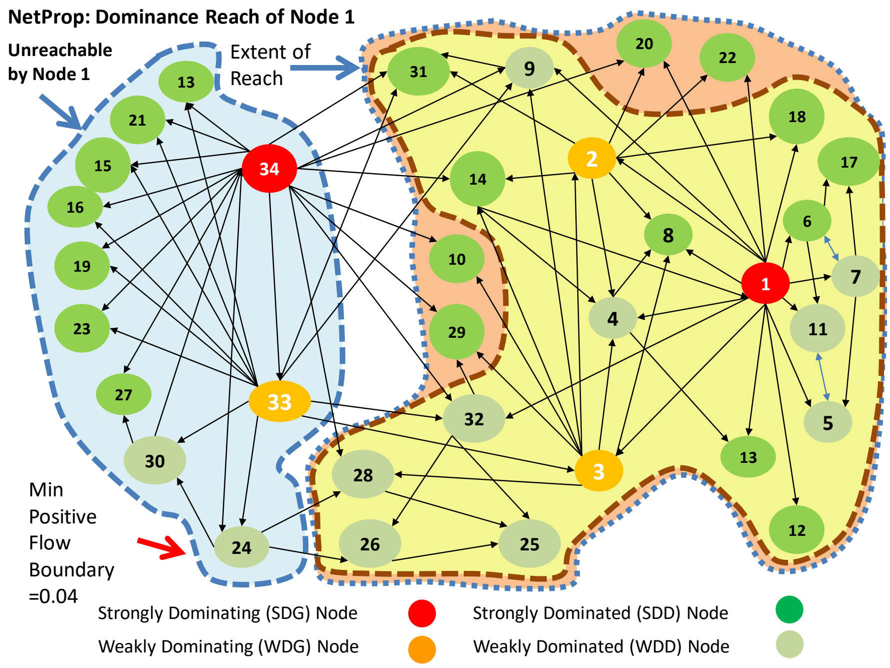


Figure A4. Network View of Propagator Community of Node 1 in ZKC Dataset

Note also that the propagator community is differently structured than NG partitions, with good reason, since NetProp considers flows and dominance, while NG partitions are modularity-based and calculated only the basis of number of edges, not the flows along them.

It is natural to ask what type of propagator community can be expected to form for Node 34 (the other dominant node). If we run NetProp seeded with Node 34, we obtain flow plot and propagator community as shown in Figure A5 and A6, respectively.

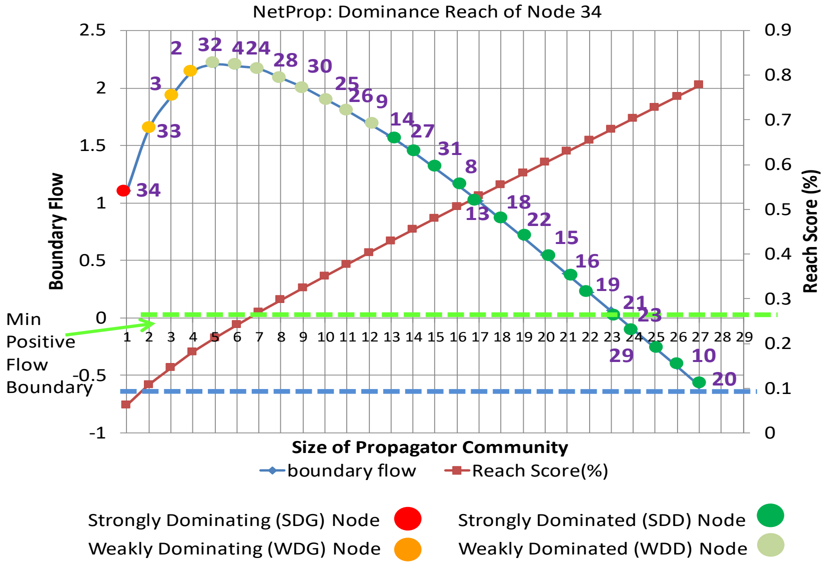


Figure A5. Boundary Flow and RS for Propagator Community around Node 34

The shape of the boundary flow curve for the propagator community around node 34 follows the familiar inverted U, as before, growing to size of 27 nodes. Minimum positive flow is 0.05, with a RS core of approximately 65%. It is evident that the dominance/influence of Node 34 penetrates deeper into the network than Node 1.

The network diagram showing the partitions is depicted in Figure A6. Here, we observe that Nodes 10 and 29, both SDDs, have links to the propagator communities of both Node 1 and Node 34. They have incoming edges from both opposing partitions, and are thus ‘sitting on the fence’, figuratively.

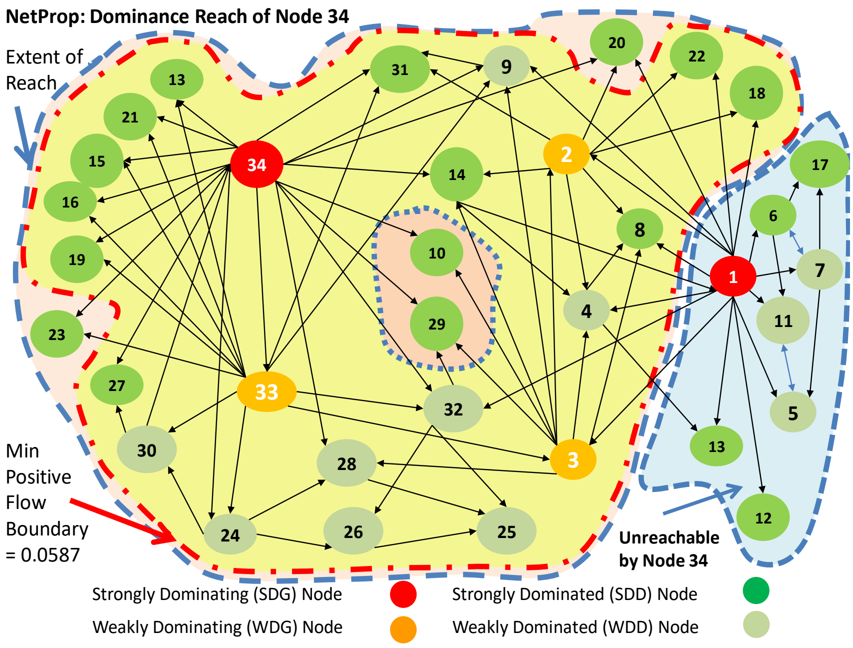


Figure A6. Boundary Flow and Reach Score with enlarging propagator community

These nodes (human agents) can perhaps be made to shift their allegiance to either community if the level of dominance by either party is enhanced. This can be demonstrated by merely changing the interaction weight representing strength of the tie. To see this, let us consider setting interaction weight of edge connecting Node 3 and Node 10, and Node 3 and Node 29, to a value of 2 instead of the default value of 1 (as for all other edges). A rerun of NetProp with these values results in the network partition of propagator community for Node 1 as shown in Figure 15.

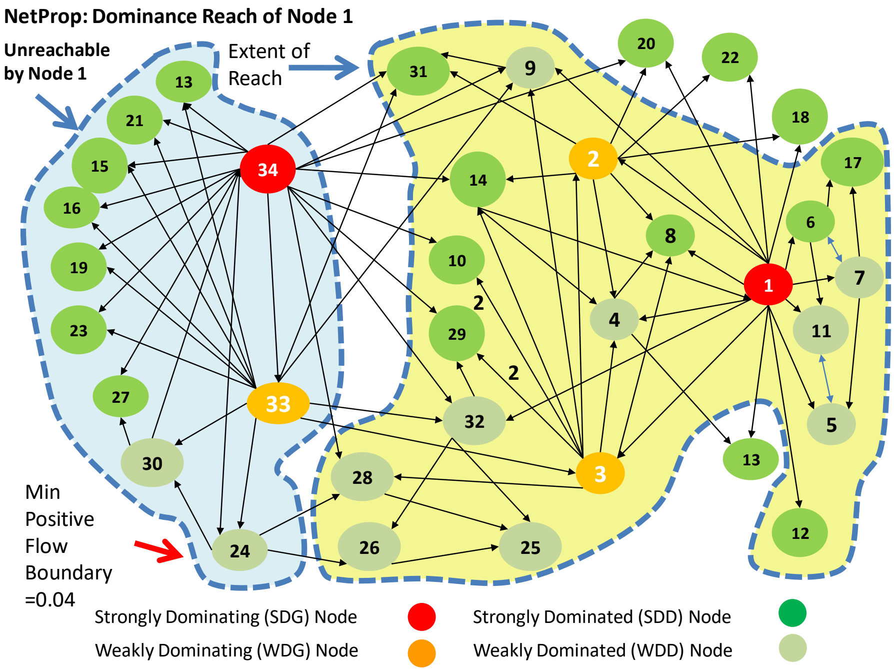


Figure A7. Minimum Posistive Boundary Flow of Propagator Community for Node 1 with weights modified for two edges, as shown

By increasing the interaction weight, note that Nodes 10 and 29, continue to be SDDs, but are now recruited earlier into the propagator community, falling inside the minimum positive flow boundary, thus playing their link role as part of Node 1's propagators. The weighting scheme lays less emphasis to the recruitment of the last four SDD nodes, 13,18,20,22.

Hence NetProp permits evaluation of spread of influence and dominance, together with reach around user selected dominant nodes in a network. An end-user of this class of NetProp results can now weigh the trade-offs that the methodology offers on whether to consider reach or dominance flow more important, and which particular nodes one should recruit for the improvement of a influence propagation goal. It also permits determination of which nodes are potential members of overlapping partitions, allowing us to calculate levels of interaction strengths which will change their membership of propagator communities. Such members may play role of weak ties who can help reach across communities.

B) Coauthorship Collaborative Network

In this example we shall examine a Co-Authorship Network, [http://konect.cc/networks/dimacs10-netscience/], which was originally published by Newman [3]. This is an undirected network of 1589 nodes representing research publication authors collaborating with colleagues on theory and experiments with networks, connected by a total of 2742 undirected edges. The number of collaborations between a pair of authors may be considered a weight on the edge representing strength of collaboration. We shall look at two scenarios: one, where edge weights are set to 1 for all edges; and two, where edge weights have non-unity normalized weights as ascribed by Newman.

Since the nodes are too many to characterize either diagrammatically, or in tabular form, we shall work with the top-10 author nodes, ranked by their DONEX scores, as tabulated in Table B1:

	Ranking with Weights set to 1			Ranking with edge weights based on number of links		
DONEX Score Rank	**Node Number**	**Name associated with Node**	**Donex Score**	**Node Number**	**Name associated with Node**	**Donex Score**
1	34	BARABASI, A	1.0627	34	BARABASI, A	1.0861
2	79	NEWMAN, M	1.0535	79	NEWMAN, M	1.0604
3	35	JEONG, H	1.0394	35	JEONG, H	1.0513
4	295	YOUNG, M	1.028	295	YOUNG, M	1.0371
5	217	BOCCALETTI, S	1.0264	282	SOLE, R	1.0333
6	282	SOLE, R	1.0255	55	OLTVAI, Z	1.032
7	55	OLTVAI, Z	1.023	217	BOCCALETTI, S	1.0285
8	63	ALON, U	1.0214	63	ALON, U	1.0281
9	220	KRUGER, T	1.0205	220	KRUGER, T	1.0275
10	97	DIAZGUILERA, A	1.0187	328	LATORA, V	1.0243

Table B1. Ranking of Authors in the Co-authorship network based on Donex Score, with unit-weights and with weights based on number links between pairs of nodes

We see that despite accounting for edge-weights characterizing numbers of linkages, there is no change in the ranking of the top4 nodes – rankings do change beyond this rank. Based on the ranking in Table B1, we shall examine the reach and propagator community around each seed node taken from the top 5donex-score ranks, with and without weights.Choosing only the top-5 makes the plots somewhat readable, and easier to interpret. The dynamic of the boundary flows, with unit-weights, and with given edge-weights, are plotted in Figure B1, and B2, respectively.

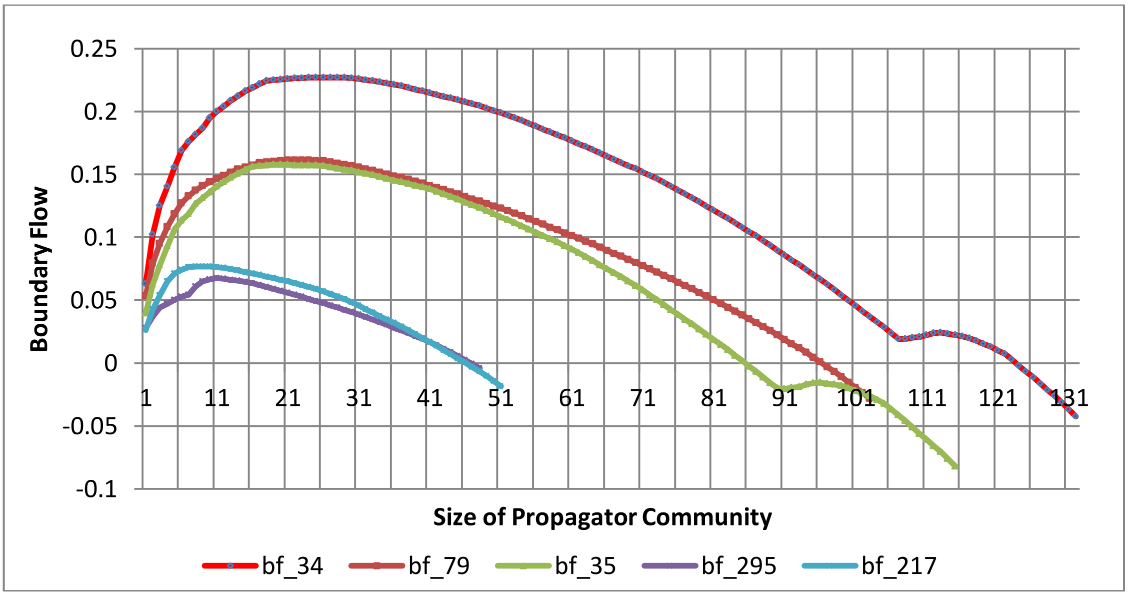


Figure B1. Boundary Flows with seeds for NetProp taken as nodes(authors) with top-5 Donex-scores. Edge-weights are taken as unity.

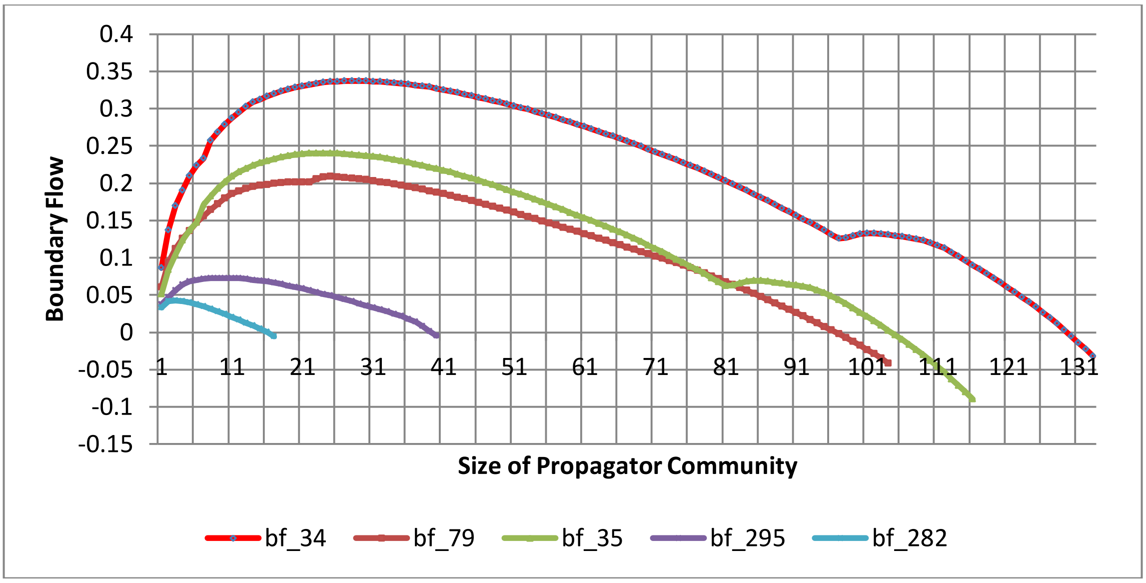


Figure B2. Boundary Flows with seeds for NetProp taken as nodes(authors) with top-5 Donex-scores. Edge-weights have non-unity strengths as specified in data.

Note that the curves have similar overall 'inverted-U' shapes, irrespective of the seed node, and follow an increasing flow until all *dominating* nodes are exhausted (maximum flow inflexion) in the ego-network around the seed, and then the flow begins to fall as *dominated* nodes are recruited into the propagator community, further until the flow turns negligibly positive; and finally negative.

We can observe firstly, that the non-unity edge-weights have the effect of increasing flows at the boundaries of sub-networks formed by the propagator communities. The largest flows in Figure B1 are about 0.1 less than the flows with non-unit weights – higher number of authorship collaborations result in higher influence flows, as expected.

The size of the propagator community around each seed node gives us a clear picture of the reach of each of the top dominant node. The 'bump' in the flow curve with author 34(Barabasi, A), and author 35(Jeong, H), at a propagator community size of about 96 and 80, respectively, is an interesting case, where a weak tie from author 35 to researcher 132 (Bianconi G), who has publication linkages to a small community of researchers working at the cusp of physics and network science. NetProp helps us identify this small, but strongly linked community of four authors - author 132 ((Bianconi G), author 204(Capocci, A), author 304(Munoz, M), and author 303(Delosrios, P). In general, such 'bumps' characterize the presence of a community around a relatively stronger dominating *local leader* node, and can be easily detected and identified with the Netprop algorithm.

References

[1] Zachary WW.,An Information Flow Model for Conflict and Fission in Small Groups, *Journal of Anthropological Research* 1977 33:4, 452-473.

[2] Newman MEJ, Girvan M, Finding and evaluating community structure in networks. Phys Rev E 69:026113.2004

[3] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices", *Phys. Rev*. E 74, 036104, May 2006.