Cooperative Computing for Autonomous Data Centers

Jared Saia (U. New Mexico)

Joint with: Jon Berry (Sandia),

Aaron Kearns (U. New Mexico)
Cindy Phillips (Sandia)

Outline

Bounding clustering coefficients

An application: Cooperative Computing

Power Law Bounded (PLB) Assumption *[Brach, et al. ’16]*

The **PLB assumption** holds for a network if, for some $\alpha > 1$:

The number of vertices with degree $\geq k$ is $O(n\, k^{1 - \alpha})$

Power Law Bounded Neighborhood (PLBN) Assumption *[Brach, et al. ’15]*

The **PLBN Assumption** holds for a network if for some α, $2 < \alpha < 3$:

For every vertex v, of degree d, the number of neighbors of v of degree $\geq$ d is: $O(d^{3 - \alpha})$

Edge Weights

Consider an edge (u,v)

Let d_u, d_v be the degrees of u, v

Let t be the number of triangles containing (u,v)

Edge Weight: $2t / (d_u + d_v - 2)$

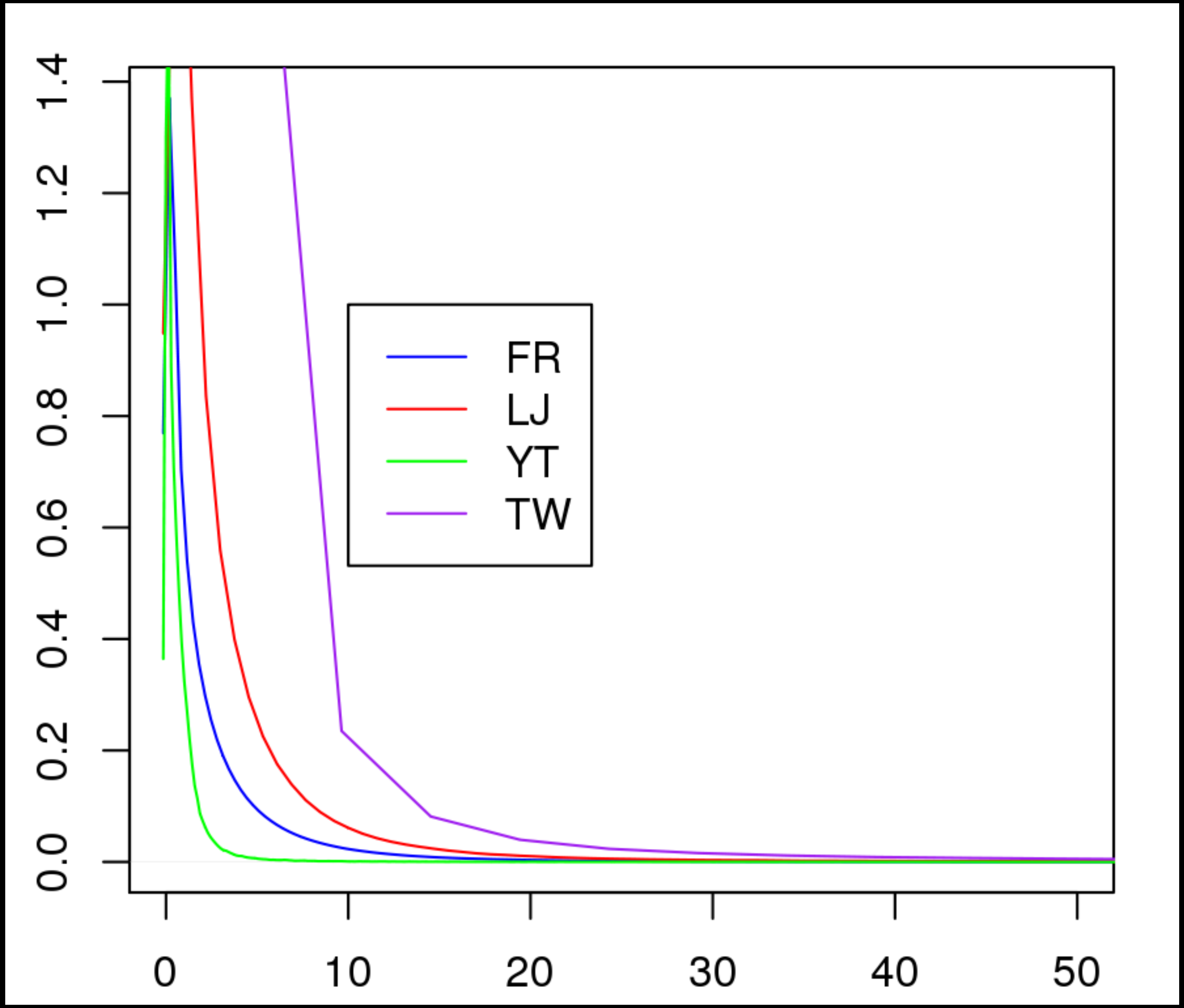
Bounded Weighted Degree (BWD) Assumption

For any human node, the sum of the weights of edges on that node is ≤ c, where c is a universal constant

Inspired by Dunbar-like argument

Weighted Degrees

Probability Density



Weighted Degree

Extremely Unlikely any vertex has weighted degree above 40

Edge Weights Revisited

Let d_u, d_v be the degrees of u, v

Let t be the number of triangles containing (u,v)

Edge Weight 1: $2t / (d_u + d_v - 2)$

Edge Weight 2: $t / (d_u + d_v - t)$

Both of these greater than:

$t / (d_u + d_v)$

$O(d^{4-\alpha})$ triangle bound

Assume:

PLBN**:** Number of neighbors of v with degree $\geq$ d is $O(d^{3-\alpha})$

BWD**:** Sum of edge weights touching v is O(1)

Then:

Number of triangles v is in is $O(d^{4-\alpha})$

Proof Sketch (1)

Neighbors with degree $\geq$ d

Each is in O(d) triangles with v

$O(d^{3 - \alpha})$ of them (by PLBN)

Contribute $O(d^{4 - \alpha})$ triangles

Proof Sketch (2)

Neighbors with degree $< d$

Let x = # of triangles these nodes contribute

Claim: Sum of edge weights is minimized when x is evenly distributed

Then, sum of edge weights is: $\Omega(d((x/d)/d)) = \Omega(x/d)$

By the BWD Assumption, this sum is $O(1)$.

Thus $x = O(d)$

Clustering Coefficient

Consider a node v, with degree d

Clustering coefficient of v:

# of triangles containing v / (d choose 2)

Prediction: clustering coefficient is $O(d^{2-\alpha})$

Cleaning Networks

Removing Non-human Nodes

Can buy 500 followers for $5 on Twitter

Non-human nodes can be densely connected

Our goal: remove non-human nodes

Removing Non-human Nodes

Our approach (sketch):

Remove nodes based on significant deviation from average weighted degree

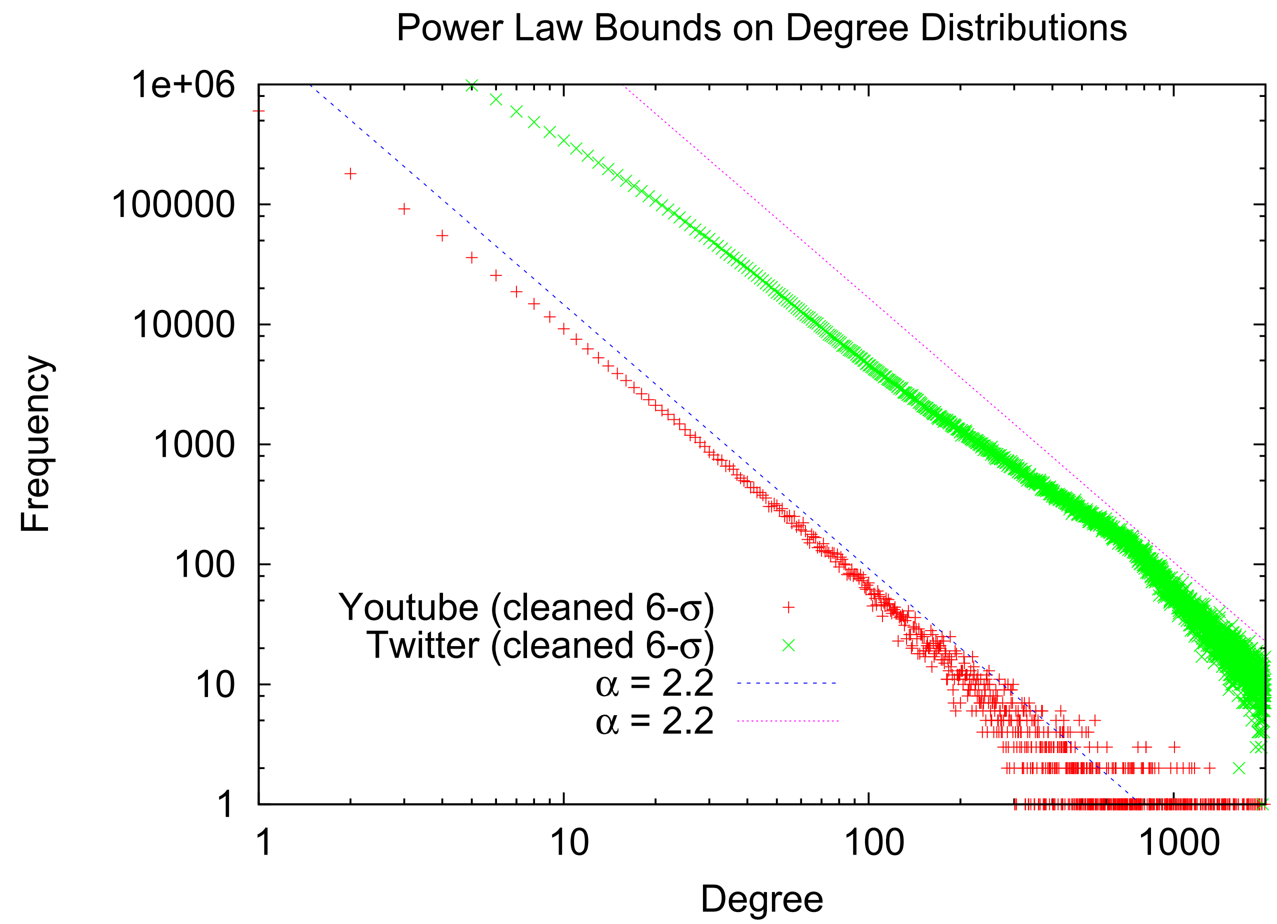
Cross-validate with human labelling and labelling based on non-topological features

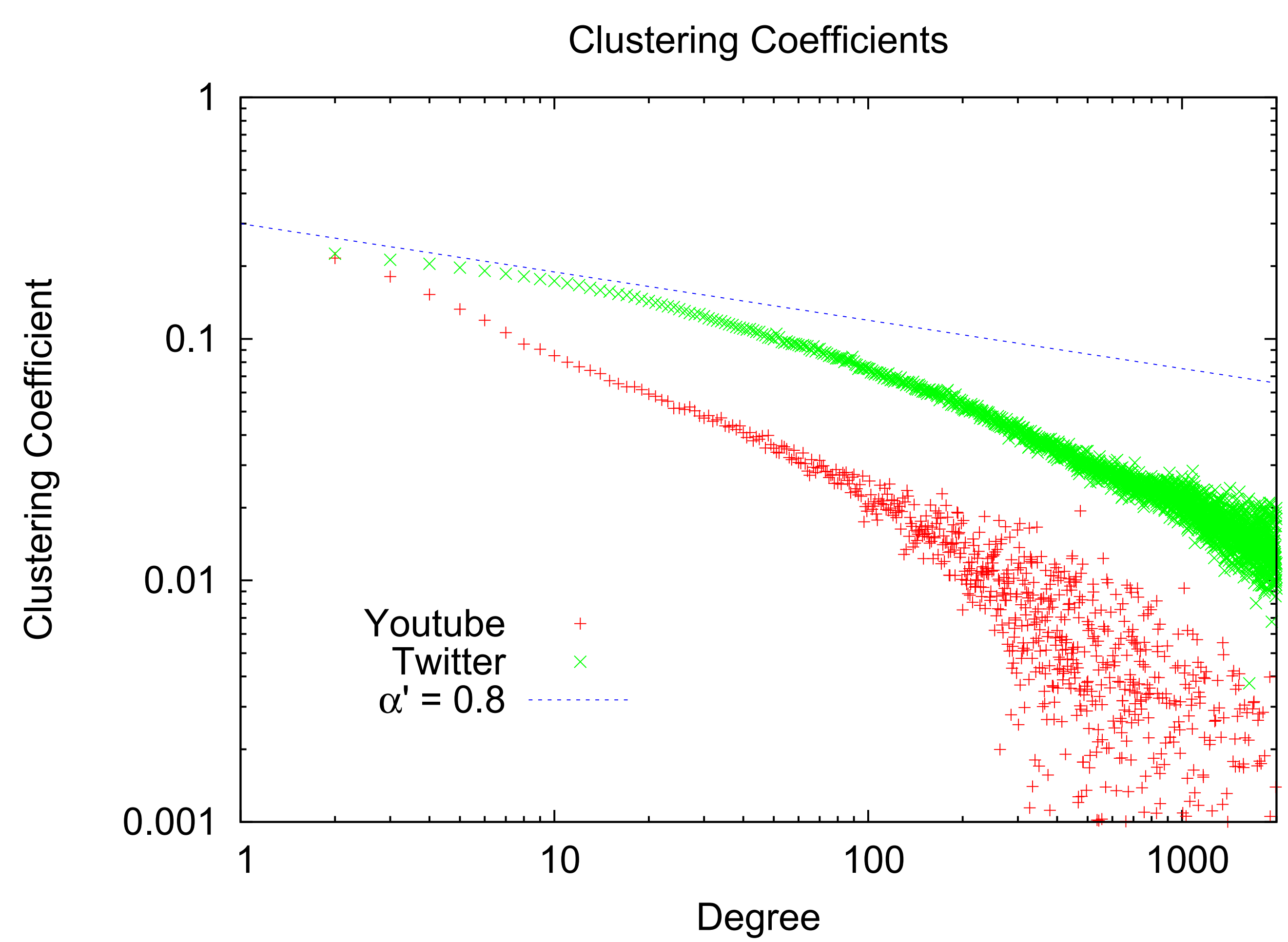
Cleaning removes few human nodes, among those we checked by hand

J. Berry, A. Kearns, C. Phillips, J. Saia. "Finding Non-human Nodes in Electronic Social Networks Using Only Topological Data" *Under Submission*, 2016.

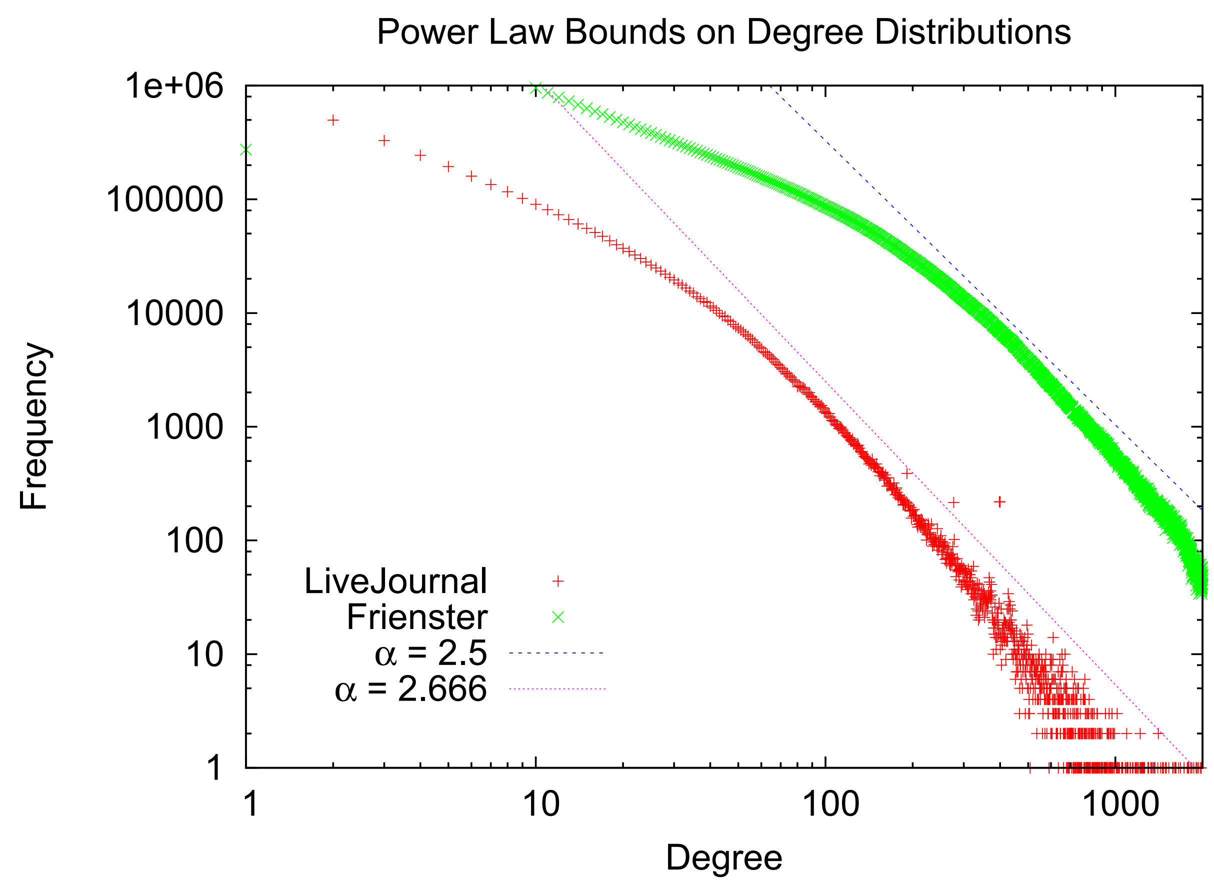
Does our Bound on Clustering Coefficients Hold on Cleaned Networks?

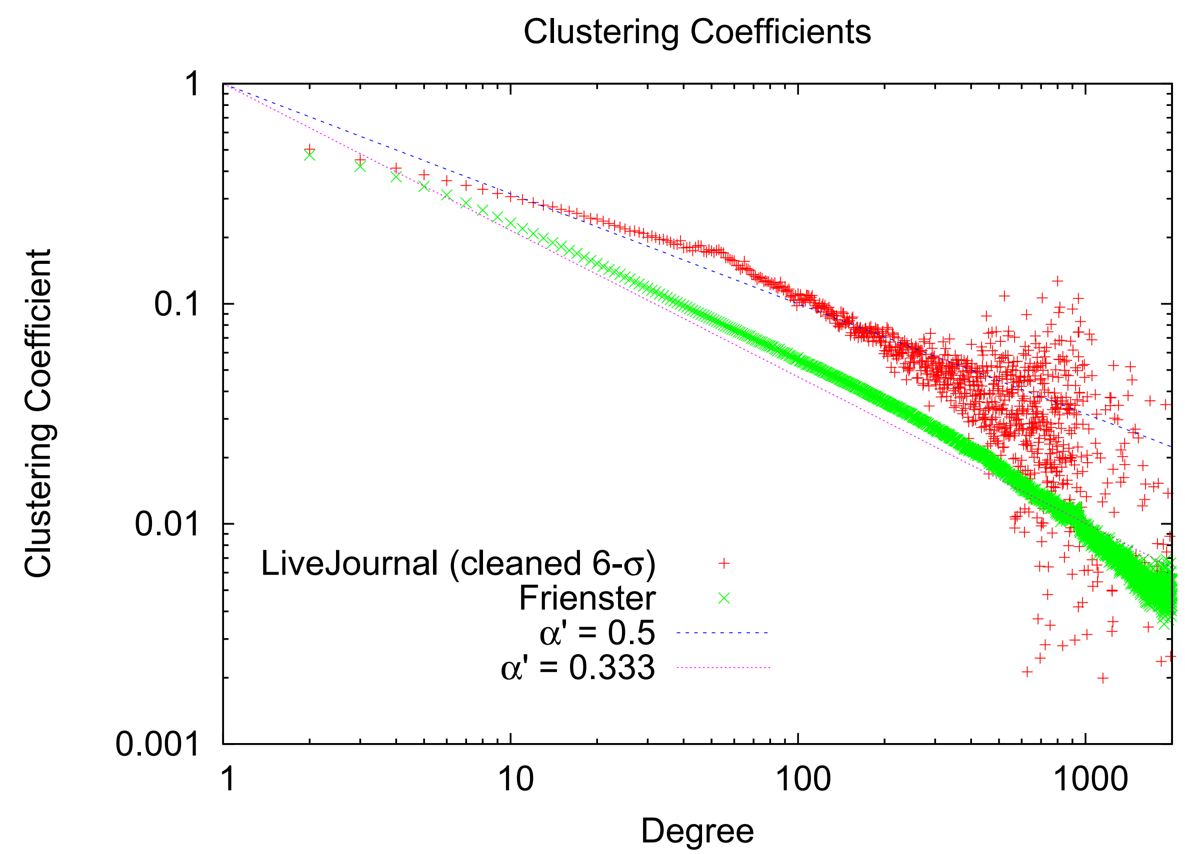
Youtube and Twitter


Power Law Bounds on Degree Distributions
Frequency
Degree
Youtube (cleaned 6-σ)
Twitter (cleaned 6-σ)
α = 2.2
α = 2.2
1e+06
100000
10000
1000
100
10
1
1
10
100
1000


Clustering Coefficients
Clustering Coefficient
Degree
Youtube
Twitter
α' = 0.8
1
0.1
0.01
0.001
1
10
100
1000

LiveJournal and Friendster


Power Law Bounds on Degree Distributions
Frequency
Degree
1e+06
100000
10000
1000
100
10
1
1
10
100
1000
LiveJournal
Friendster
α = 2.5
α = 2.666


Clustering Coefficients
Clustering Coefficient
Degree
1
0.1
0.01
0.001
1
10
100
1000
LiveJournal (cleaned 6-σ)
Friendster
α' = 0.5
α' = 0.333

The Takeaway

Preliminary experiments suggest clustering coefficient is $O(d^{2-\alpha})$

Hidden constant seems fairly small

Not necessarily tight

Only holds after cleaning

Our Application:
Cooperative Computing

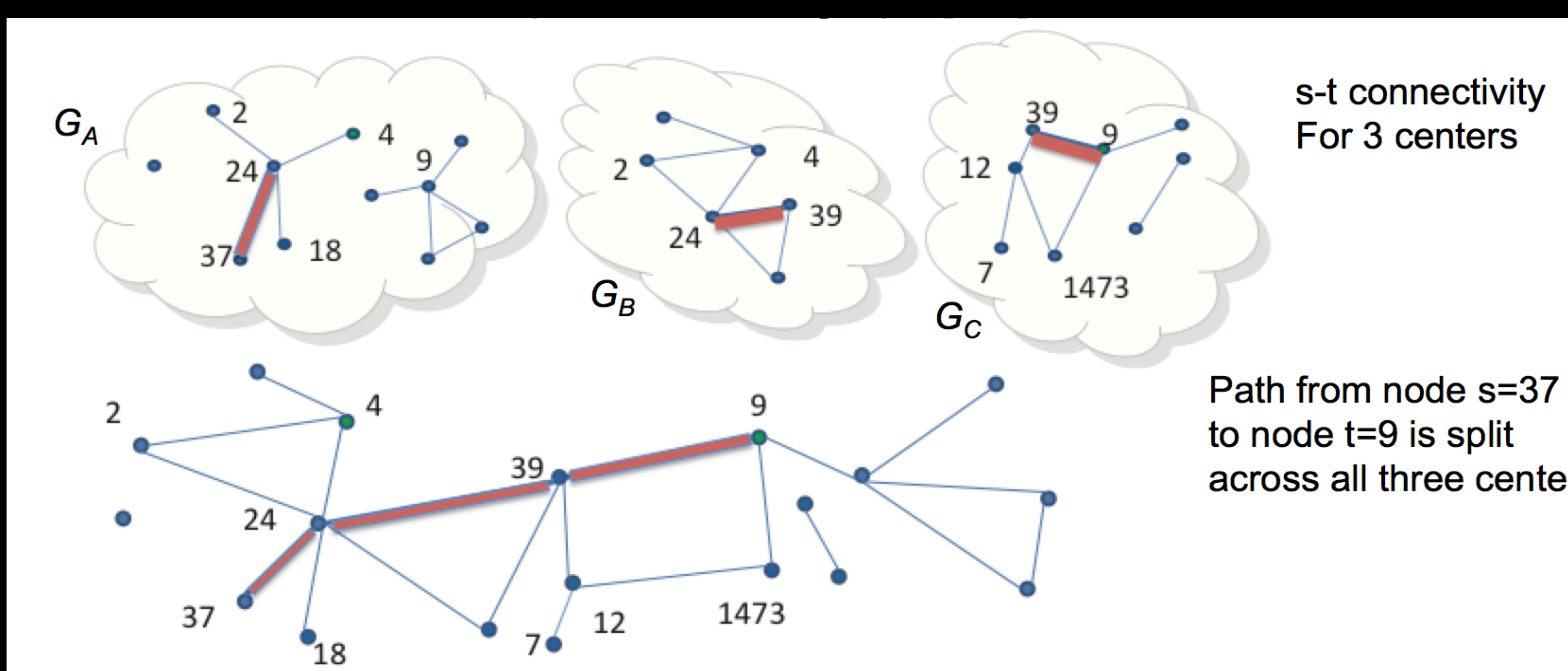
Cooperative Computing

Alice and Bob (or more) have graphs G_A and G_B

- Share the name space for nodes

Cooperate to solve problems on $G_A \cup G_B$

- Goal: limit total communication



Previous Work: Algorithms for s-t connectivity

Assumes social network structure: giant component

$O(\log^2 n)$ communication cost

Can ensure players get no information beyond the output in *honest-but-curious* model

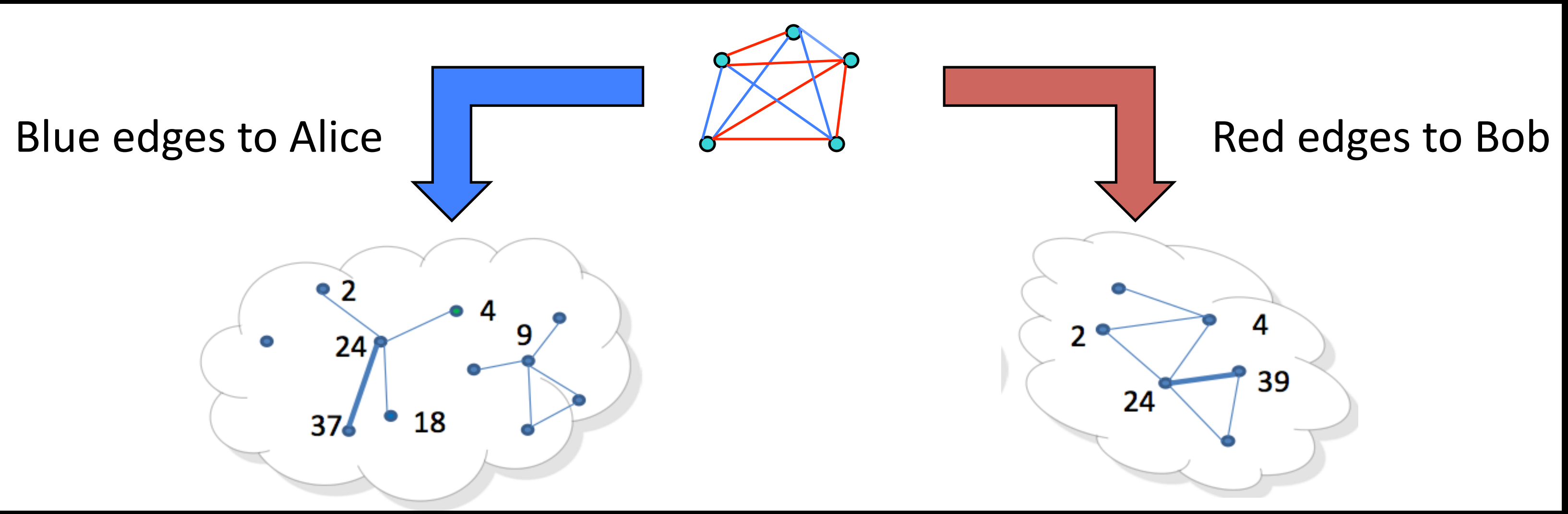
J. Berry, J, M. Collins, A. Kearns, C. Phillips, J. Saia, and R. Smith. "Cooperative Computing for Autonomous Data Centers." *IEEE Parallel and Distributed Processing Symposium (IPDPS)*, 2015.

Finding a clique in $G_A \cup G_B$

Adversary plants a clique of say polylog nodes

Assigns the edges to Alice and Bob

Can we find the clique with little communication?

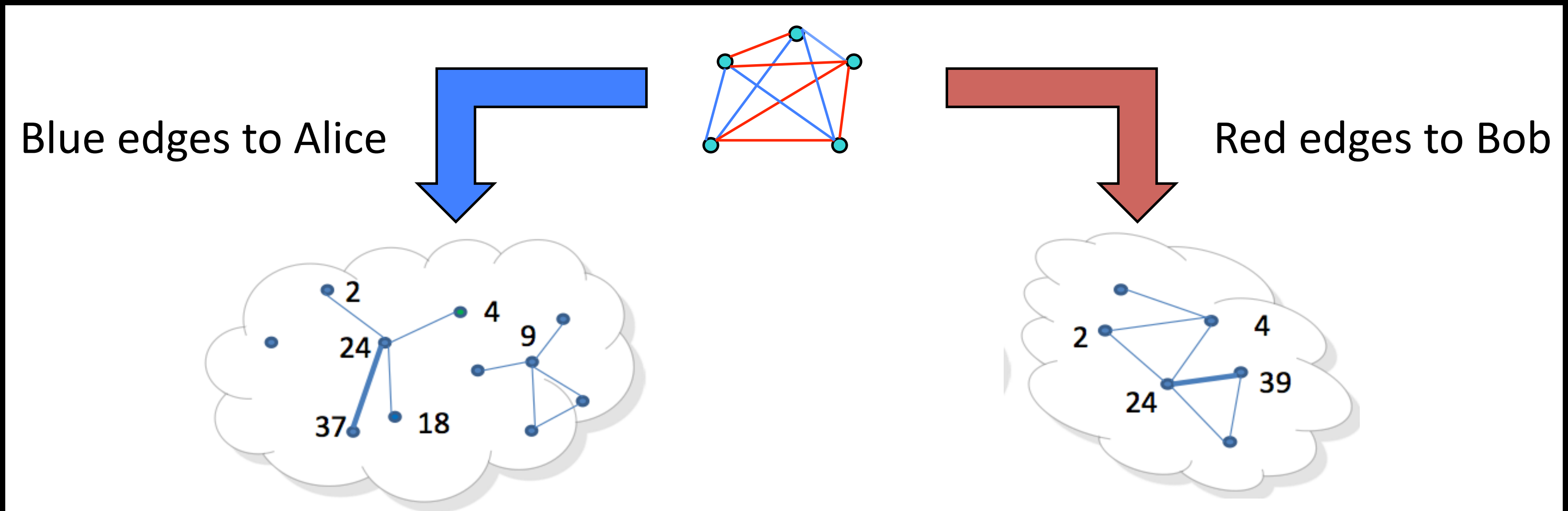


Finding a clique in $G_A \cup G_B$

Q: Can social network structure help us?

A: Bound the number of naturally occurring triangles

Decreases communication costs



Questions?

References

Edge Weight 1: [J. Berry, B. Hendrickson, R. LaViolette, and C. Phillips, “Tolerating the community detection resolution limit with edge weighting,” Physical Review E, Vol. 83, No. 5, 2011.]

Edge Weight 2: [Easley, David, and Jon Kleinberg. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, 2010.]

References

PLB and PLBN Assumption: [Brach, P., Cygan, M., Łącki, J. and Sankowski, P., Algorithmic Complexity of Power Law Networks, Symposium on Discrete Algorithms (SODA), 2016.]