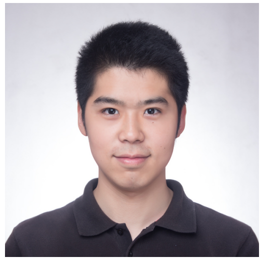
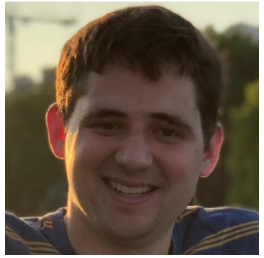
The Closure Coefficients:

A New Perspective on Network Clustering

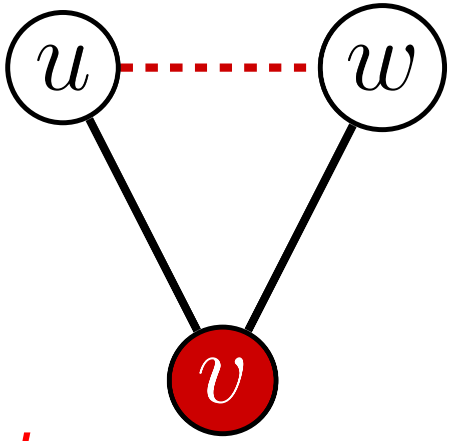

Hao Yin
Stanford University
yinh@stanford.edu

Austin R. Benson
Cornell University
arb@cs.cornell.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

Background: clustering phenomenon

- Observation: An increased chance of edge existence between nodes with a common neighbor (aka, triadic closure)



- Metric: the clustering coefficient

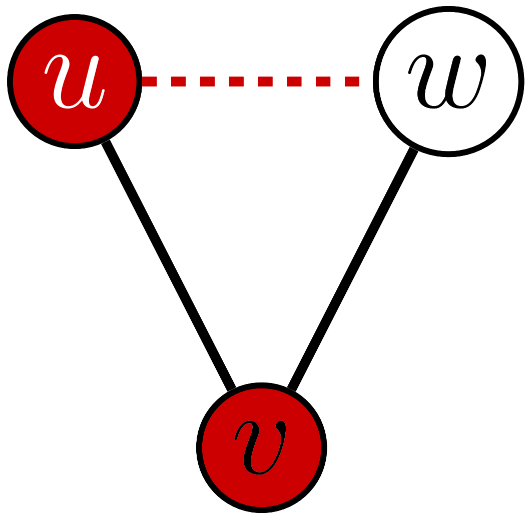
  $C(v) = \frac{2 \cdot T(v)}{d_v(d_v - 1)}$, fraction of length-2 path *centered* at node v that are closed

- Used in
  - Role discovery [Henderson et al. 2012, Ahmed et al. 2018]
  - Outlier detection [LaFond et al. 2014]
  - Psychology [Bearman et al. 2004]

Background: clustering explanation

- Explained by local evolutionary processes:
  - Social friendship network
  - Citation
- Question: which node closes this length-2 path?

  center? head?



- A fundamental gap in network science between how clustering is ***measured*** and ***explained***!

Outline

- Propose a new and simple metric of triadic closure which is based on the head node, the *closure coefficient*.
- Theoretical and empirical properties:
  - popular nodes are more likely to close triangles;
  - useful theoretical tool in graph analysis, e.g., community detection;
  - correlation with temporal triadic closure.

Definition: closure coefficient

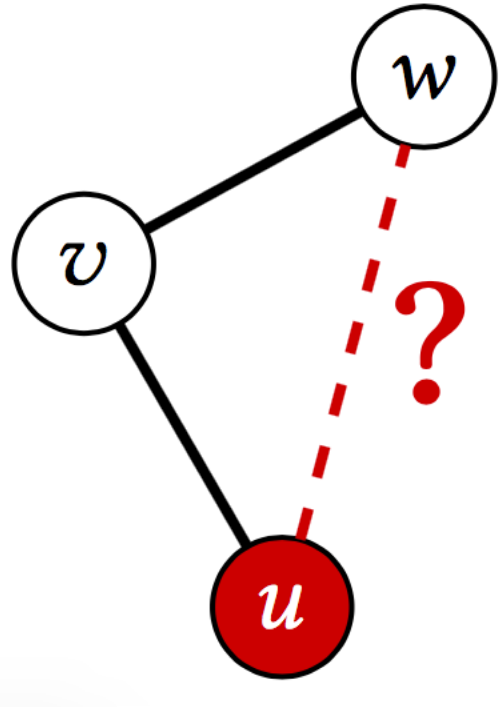
- (Local) closure coefficient: fraction of length-2 path headed at node u that are closed:

$$\mathrm{H}(u) = \frac{2 \cdot T(u)}{W(u)}$$

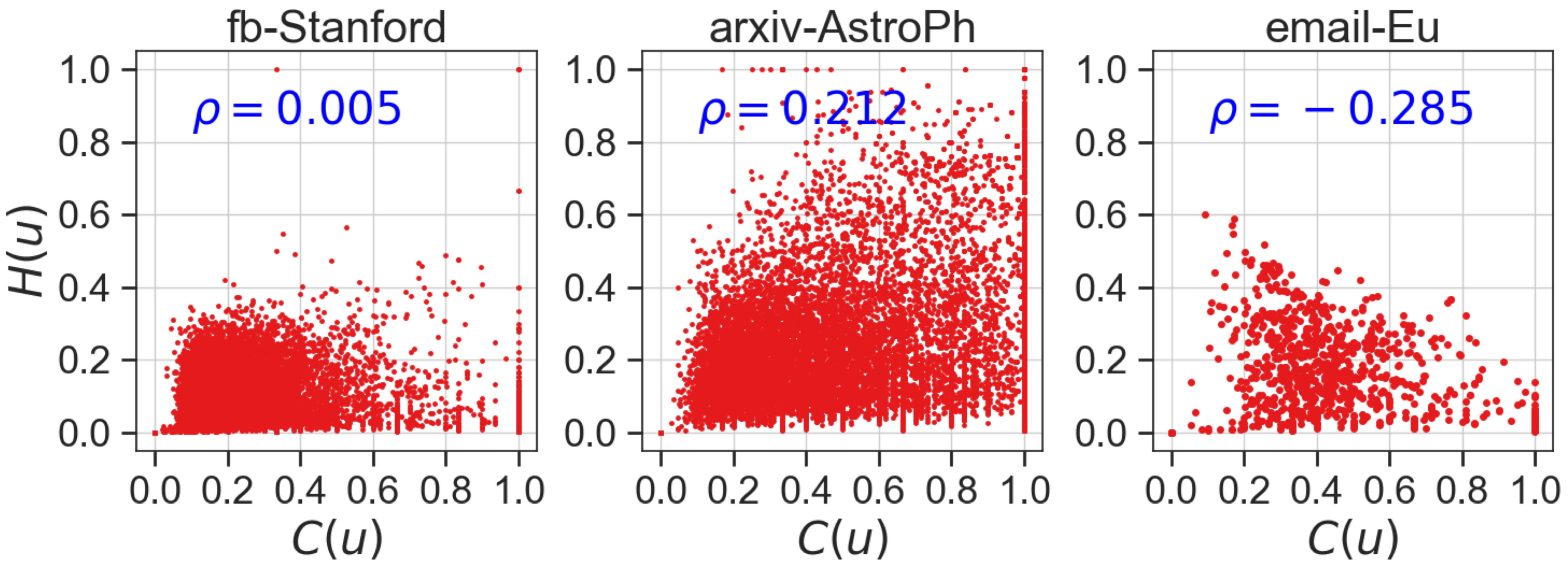
- Computation

$$W(u) = \Sigma_{v \in N(u)}(d_v - 1) = \Sigma_{v \in N(u)} d_v - d_u$$

✓ Requires the same computational effort as the clustering coefficient!

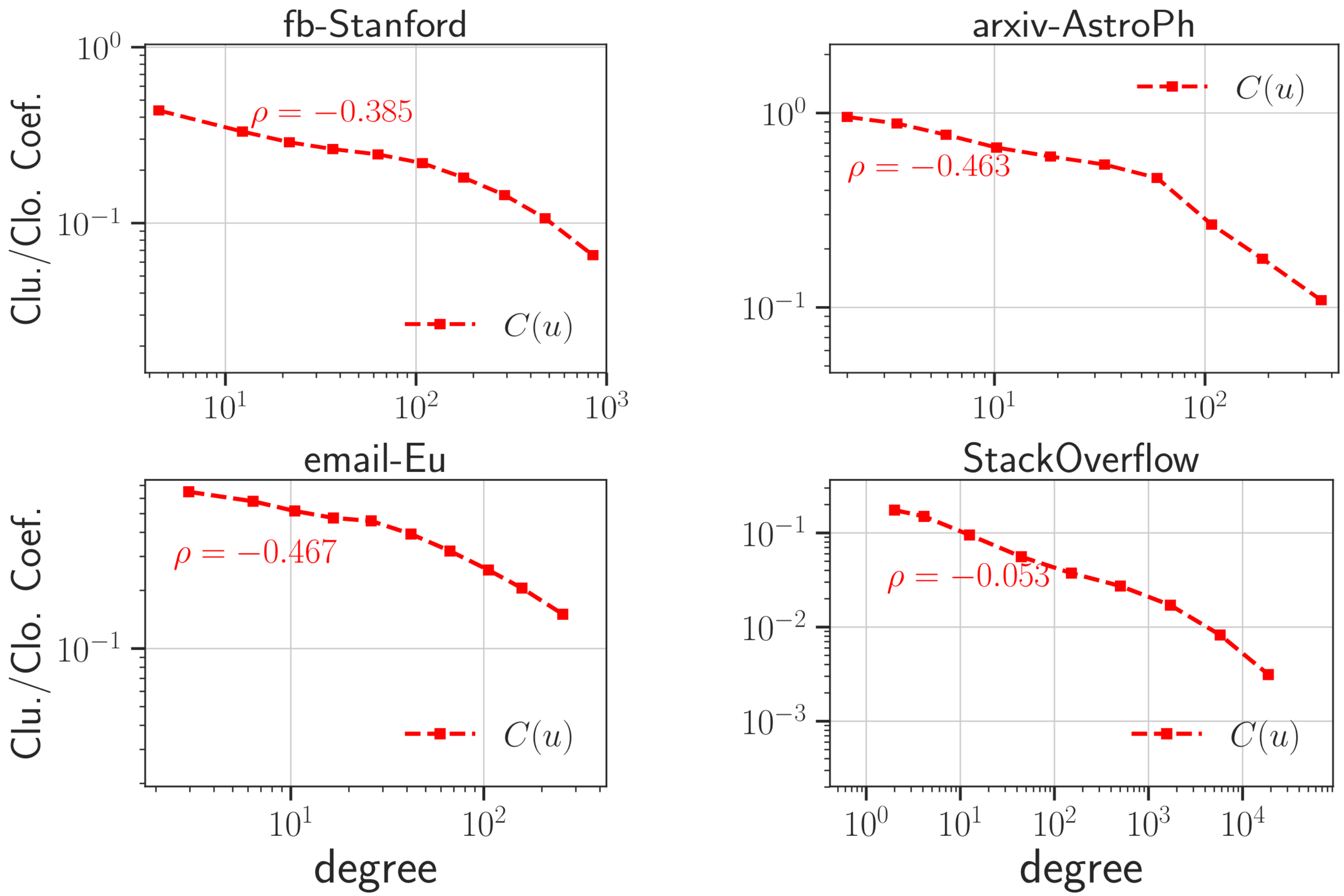


Definition: closure coefficient

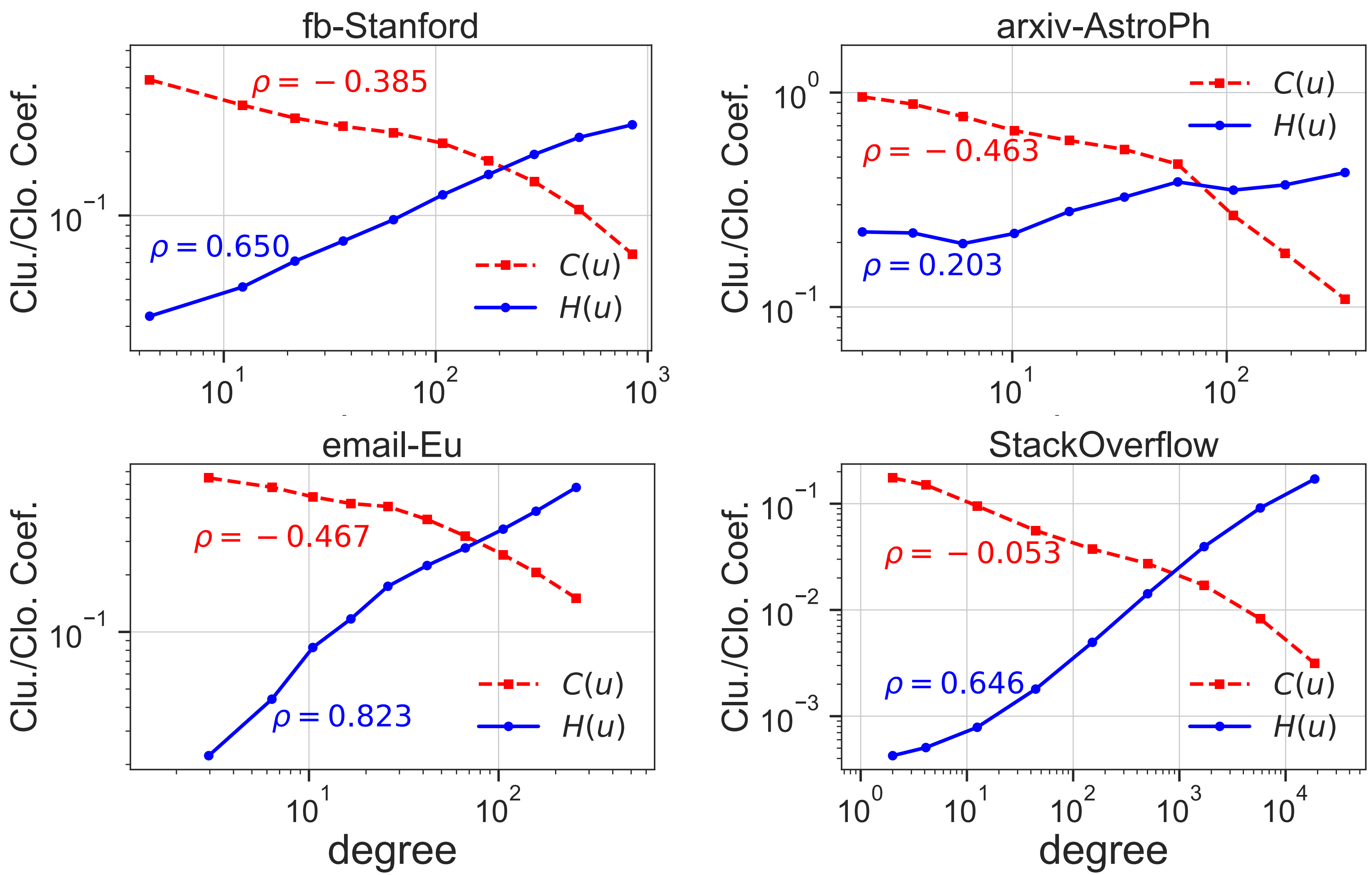


Weak correlation between clustering and closure coefficient!

Property: increase with node degree



Property: increase with node degree



Property: increase with node degree

[Background] The Configuration Model

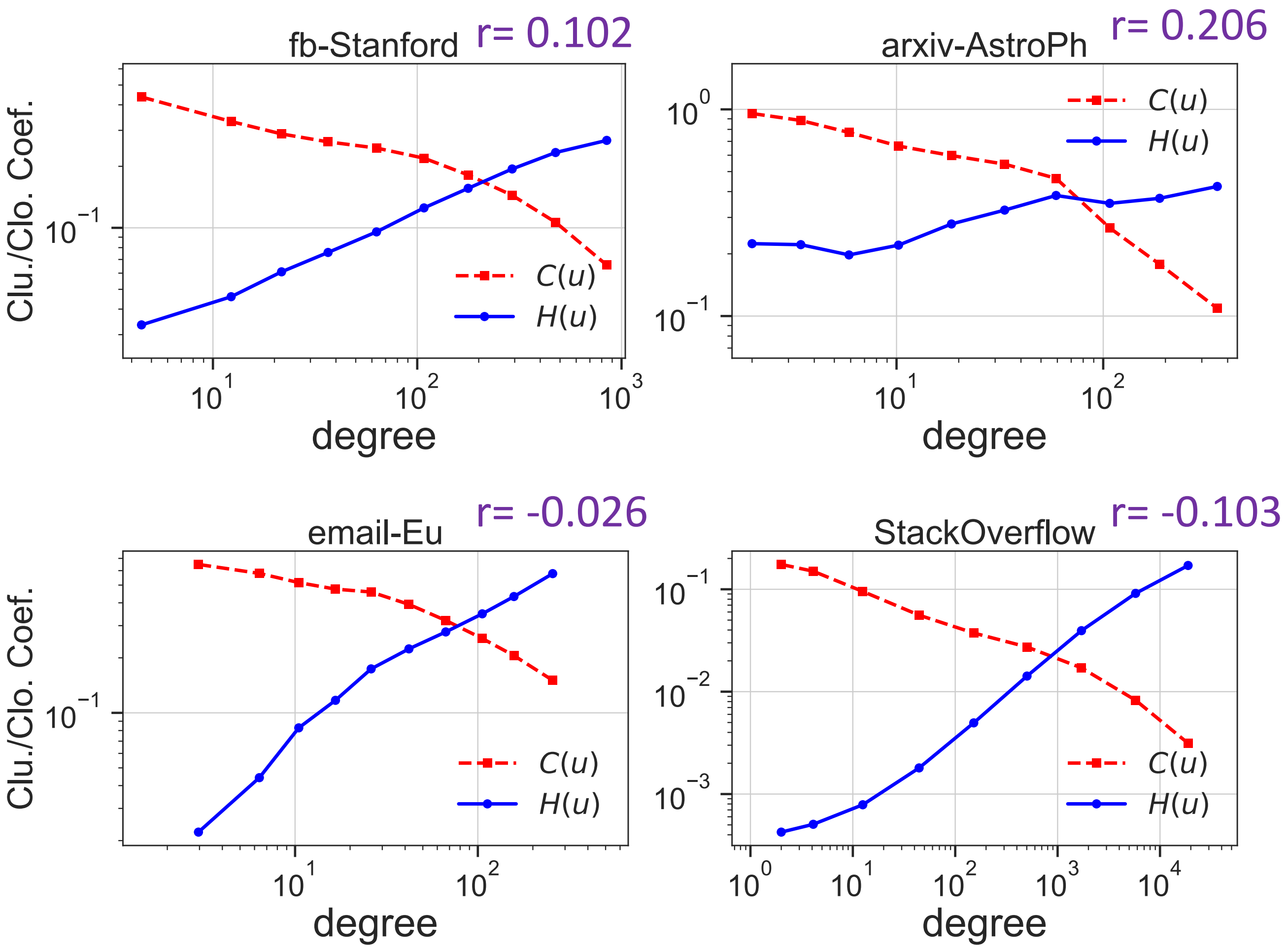
- A uniform distribution over all graphs with a specified degree sequence (distribution).

[Theory] In the configuration model with any degree distribution $\{p_k\}_{k=1}^{\infty}$, the closure coefficient at any node u satisfies

$$\mathbb{E}[H(u)] \sim (d_u - 1) \cdot const$$

as $n \to \infty$.

Property: increase with node degree



$$\mathrm{H}(u) = \frac{2 \cdot T(u)}{\sum_{v \in N(u)} d_v \; - d_u}$$

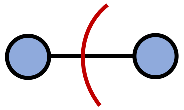
- $\log \mathbb{E}[H(u)] \approx 1 \cdot \log d_u + const$
- In practice, the increase is slower than that under configuration model.
- Can be partly explained with degree assortativity.
- An open problem.

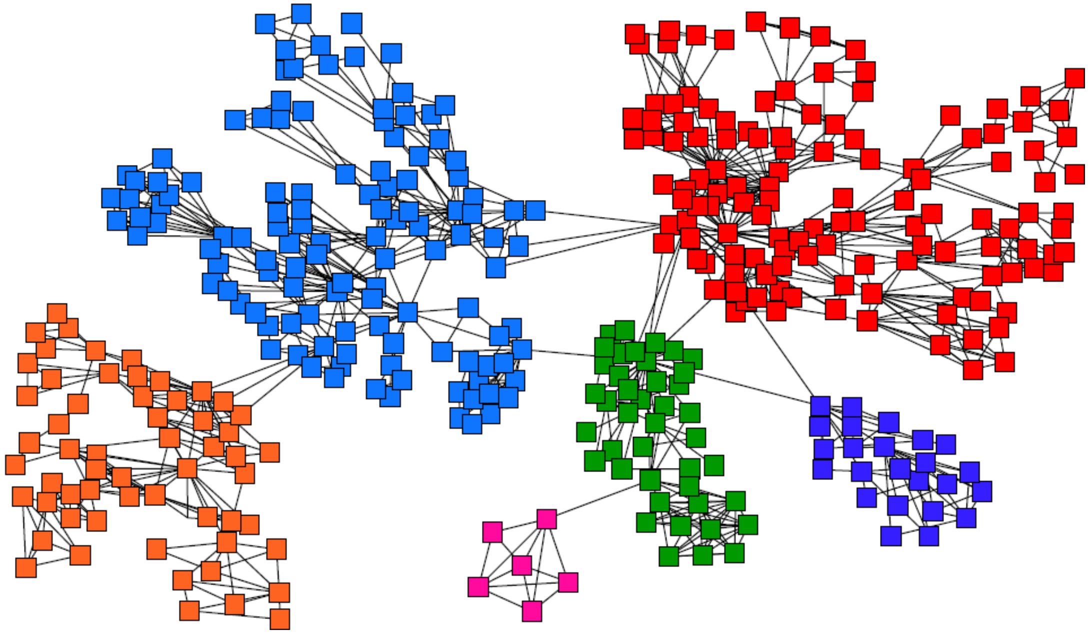
Property: community detection

Background: community structure and detection

Usually formulated as finding a set of nodes S with minimal conductance [Schaeffer, 07].

$$\phi(S) = \frac{\#(edges\ cut)}{Vol(S)}$$

- $edges\ cut$: 
- $Vol(S) = $ #(edge end points in S)

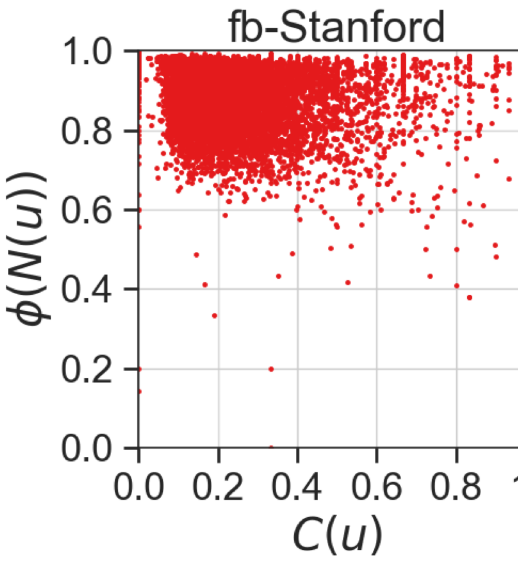


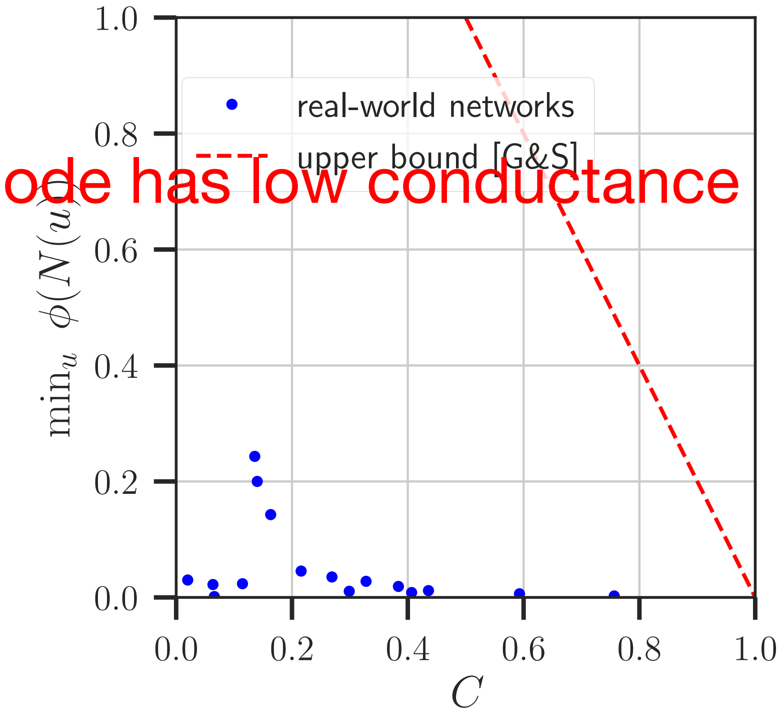
Property: community detection

[Background] clustering coefficient and community detection

- [Gleich & Seshadhri 12]: For any network, there exists a node u^* s.t. $\phi\big(N(u^*)\big) \leq 2(1 - C)$ where C is the global clustering coefficient.

  - This upper bound is very loose.
  - We have little information on which node has low conductance neighborhood.





$C(u)$

Property: community detection

- [Gleich & Seshadhri 2012]: There exists a node u^* such that $\phi\left(N(u^*)\right) \leq 2(1 - C)$.
- [Our result]: For any node u, we have $\phi\left(N(u)\right) \leq 1 - H(u)$.

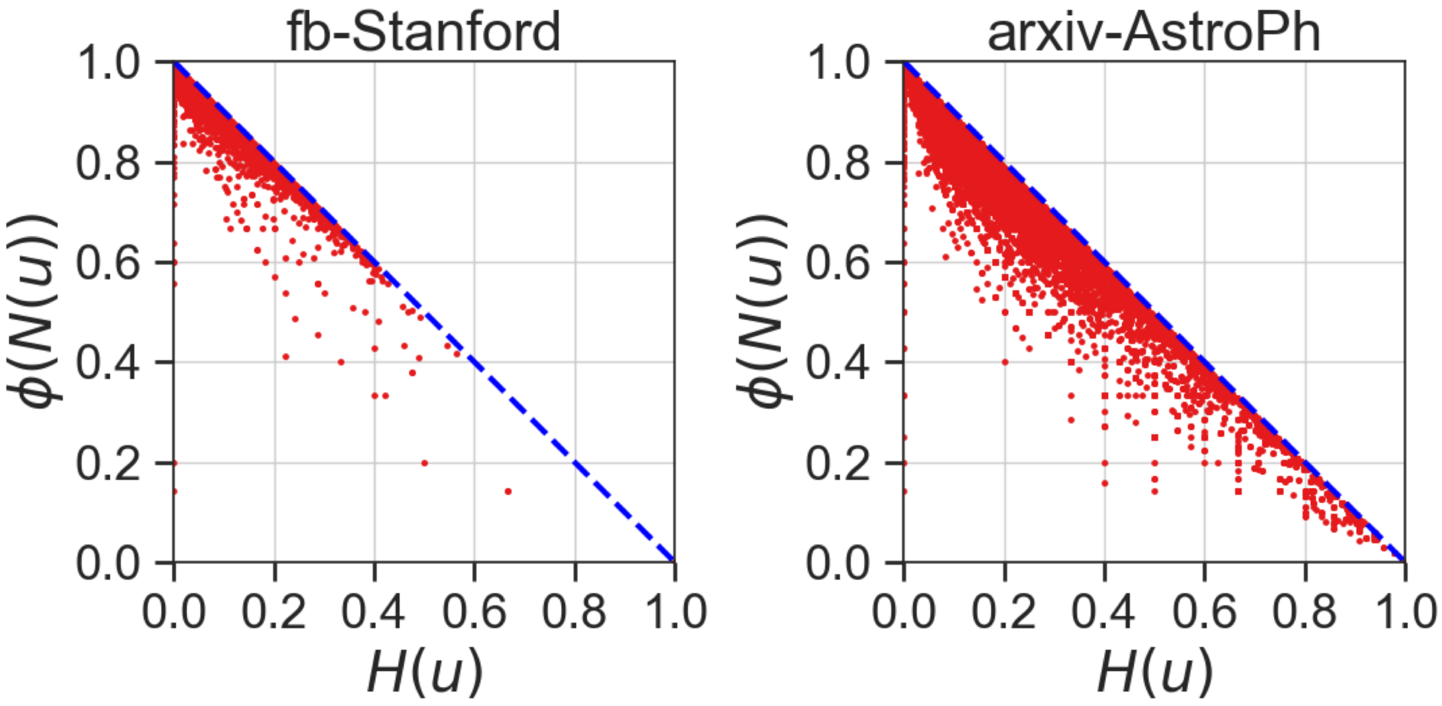
Stronger result

- Applies to every node, not only the best one;
- Gives a tighter upper bound on the best neighborhood conductance
  - note: $\max_{u} H(u) \geq C$, since the global clustering coefficient is a weighted average of closure coefficients.

A simple and elegant proof

Property: community detection

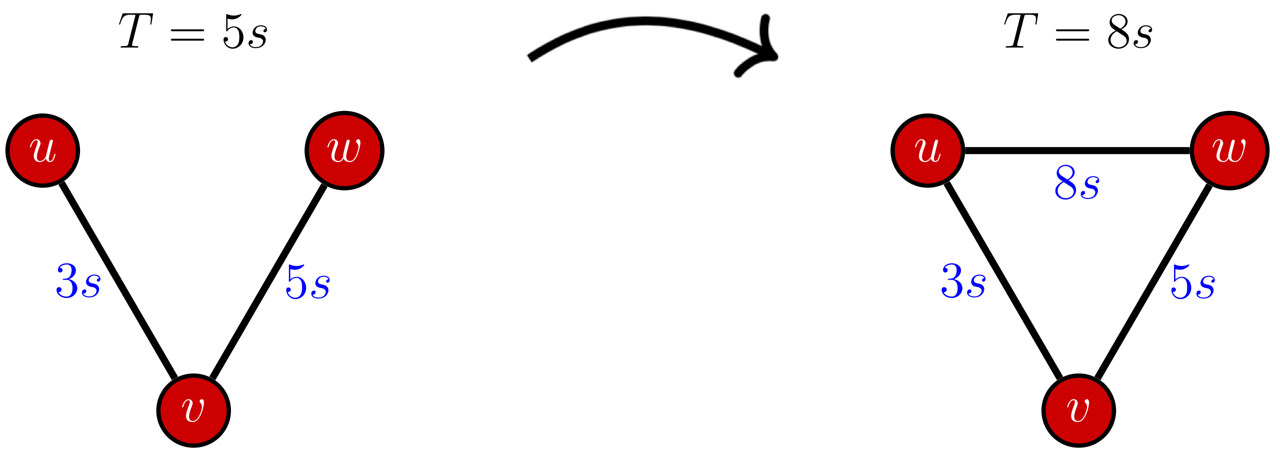
- [Our result]: For any node u, we have $\phi(N(u)) \leq 1 - H(u)$.



Property: temporal triadic closure

- Clustering and closure coefficients measure the rate of triadic closure on static network.
- However, triadic closure is a graph dynamic evolutionary process.

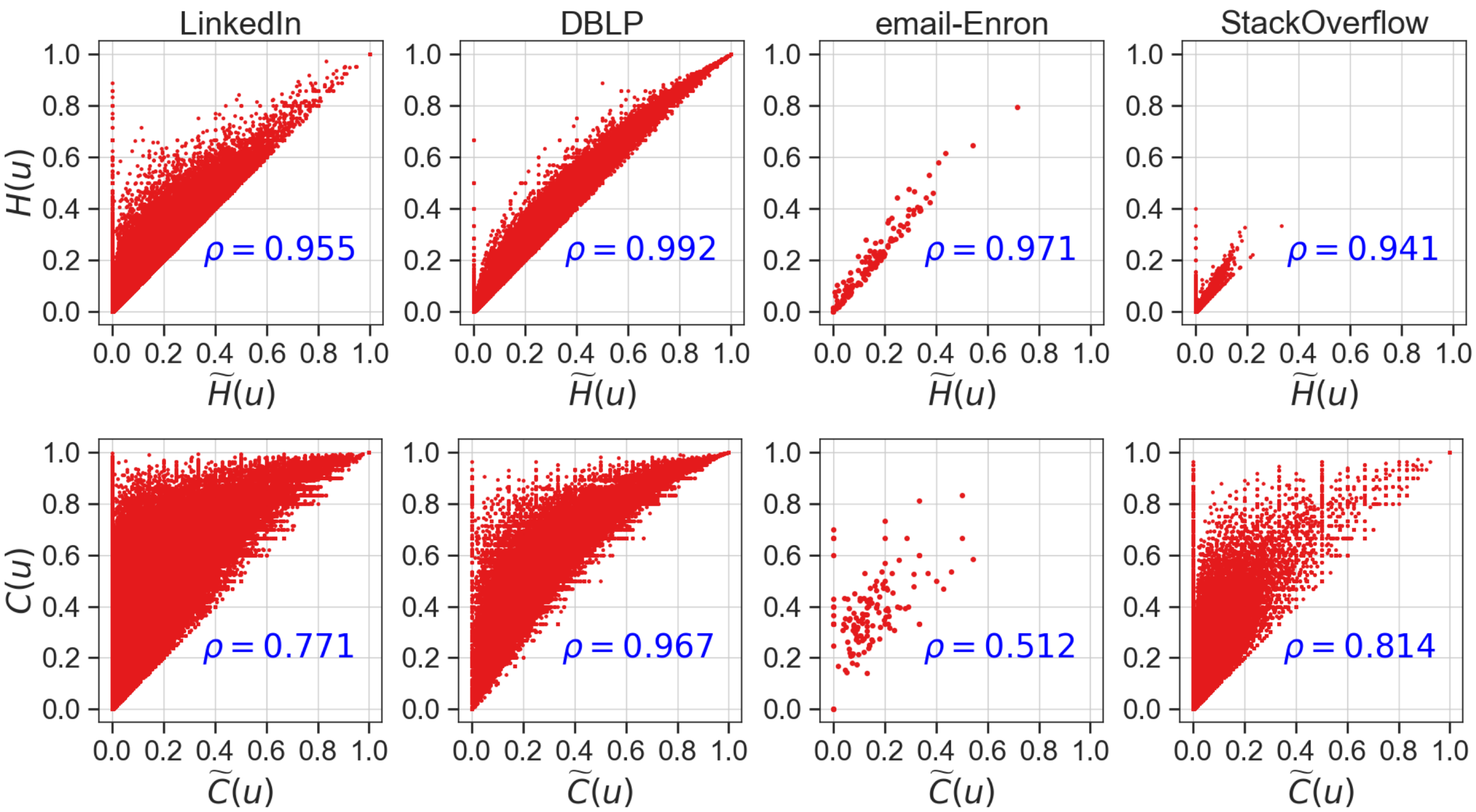
Property: temporal triadic closure



Definition

- temporal wedge
- temporal closure coefficient $\widetilde{H}(u)$
- temporal clustering coefficient $\tilde{C}(v)$

Property: temporal triadic closure



Recap

- Introduced the local closure coefficient
  - A simple metric for head-based local clustering
- Theoretical and empirical properties
  - increase with node degree
  - useful theoretical tool in graph analysis
  - correlation with temporal triadic closure

Future work

- Applications in network-based machine learning problems
- Theory for the correlation with temporal closure coefficient

Thanks!

Hao Yin
Stanford University
yinh@stanford.edu

Collaborators

Austin R. Benson
Jure Leskovec

Details are found in

- Hao Yin, Austin R. Benson, Jure Leskovec. The local closure coefficients: a new perspective on network clustering. In *WSDM,* 2019.