Evaluation of Community Similarity Based on Hierarchical Distance

Ricky Laishram
rlaishra@syr.edu

Sucheta Soundarajan
susounda@syr.edu

Syracuse University,
Syracuse, NY, USA

NERCCS 2018

Motivation

Commonly used measures to evaluate community similarity:

- Average Jaccard Similarity
- Normalized Mutual Information

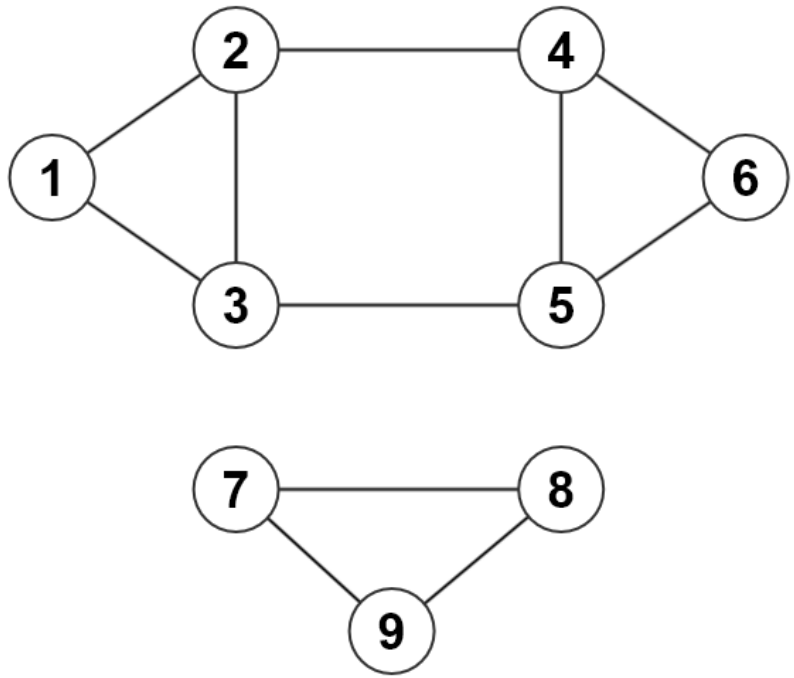


Figure: Graph G

Motivation

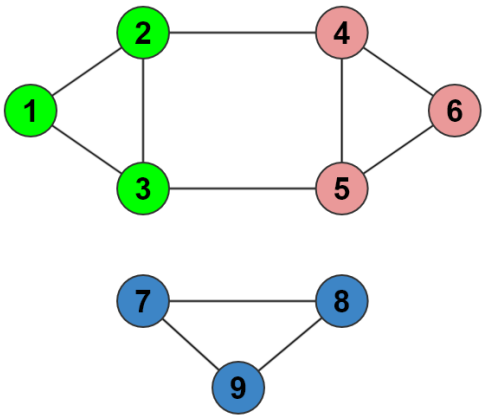


Figure: Partition P

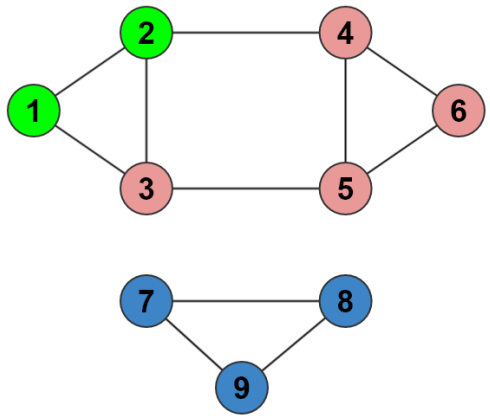


Figure: Partition P'

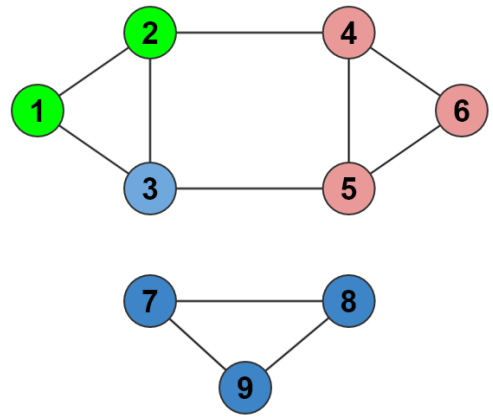


Figure: Partition P''

Motivation

$$P = \{\{1,2,3\},\{4,5,6\},\{7,8,9\}\}$$
$$P' = \{\{1,2\},\{3,4,5,6\},\{7,8,9\}\}$$
$$P'' = \{\{1,2\},\{4,5,6\},\{3,7,8,9\}\}$$

$$J\left(P,P'\right) = J\left(P,P'\right)$$
$$NMI\left(P,P'\right) = NMI\left(P,P''\right)$$

Set partition similarity measures are not very good measures of community partition similarity.

Our Goals

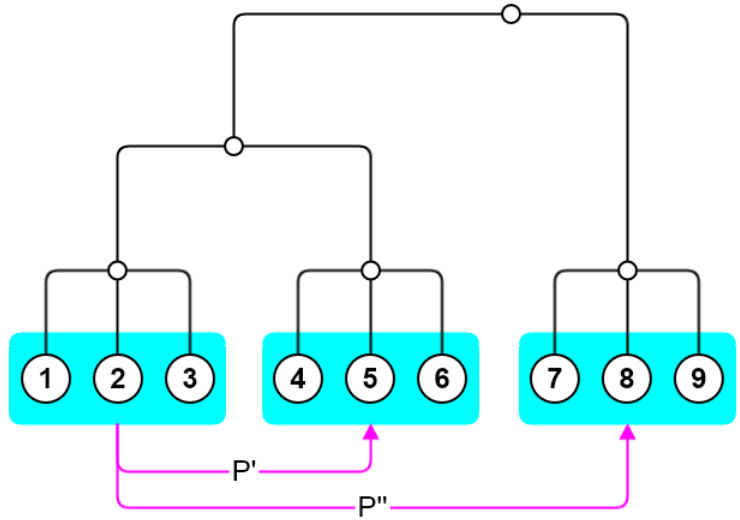
- Axiomatic approach to community similarity measures.
- Account for the hierarchical community structure.
- Design similarity measure that takes into account these axioms.

Axion 1: Community Closeness

$\psi^G(A, B)$: Community partition similarity between A, B in graph G.

P: Original Partition

P', P'': Modified Partitions



$$\psi^G(P, P') > \psi^G(P, P'')$$

Axiom 2: Sample Size Monotonicity

V_α: Subset of V of size α.
P_{V_α}: Partition of V_α with P.

For $\alpha > \beta$,

$$\mathbb{E}_{V_\alpha}\left[\psi^G\left(P, P'_{V_\alpha}\right)\right] \geq \mathbb{E}_{V_\beta}\left[\psi^G\left(P, P'_{V_\beta}\right)\right]$$

If the equality holds: Sample Size Independence.

Hierarchical Community Similarity

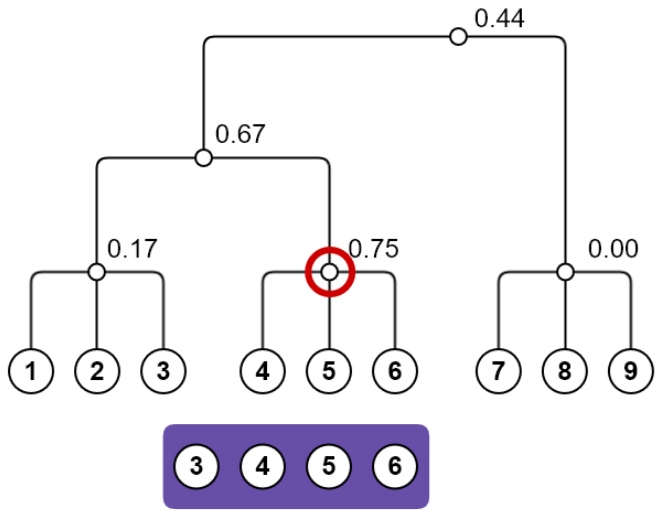
1. Assign Height based on some quality measure Q^G.
2. Match each community in P, P' to one in the dendrogram.
3. Compute Hierarchical Closeness.

Community Matching

Match $p_0 \in P'$ to a community in the dendrogram.

Any measure of set similarity can be used.

In the example, Jaccard Similarity is used.



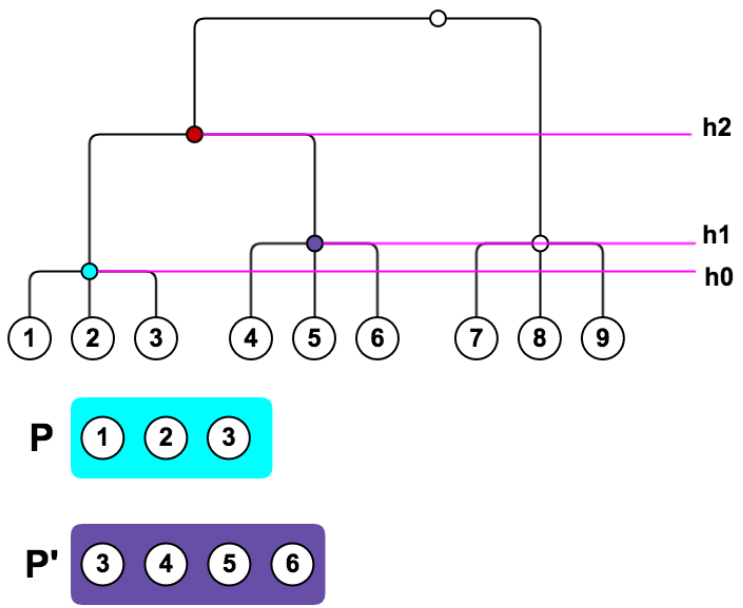
Hierarchical Closeness

Closeness of node u_3:

$$\mathcal{T}_{P,P'}(u_3) = 1 - \left(h2 - \frac{1}{2}(h1 + h0) \right)$$

Compute the hierarchical closeness for all nodes $u \in V$.

Finally, compute the mean of all the hierarchical closeness to get the Hierarchical Community Similarity (HCS) between the two partitions.



Experiment: Community Closeness

Experimental Setup:

- Get some partition P of V.
  - In our experiment, P is the partition that has maximum modularity.
- Generate P', P'', as described before.
- Check if *Community Closeness* is satisfied, for the different measures.
- Repeat k times. (In our experiment $k = 30$.)

Experiment: Community Closeness

Table: Percentage of trials, in which *Community Closeness* is satisfied.

Dataset	HCS	rHCS	Jaccard	F1	NMI	rNMI	FCV	HSD	hNMI
Fb_Amherst	**100**	**100**	20	67	80	73	00	43	37
Ca_Erdos	**100**	**100**	13	63	87	57	00	43	47
Soc_Hamsterster	**100**	**100**	00	97	**100**	63	00	40	60
Inf_USAir	**100**	**100**	03	97	97	67	00	70	53
Bio_Yeast	**100**	**100**	43	33	57	47	00	57	67
Aggregate	**100**	**100**	16	71	84	61	00	51	53

HCS and rHCS satisfies *Community Closeness* in every experiments.

Experiment: Sample Size Monotonicity

Experimental Setup:

1. Get partitions P, P' of V.
2. Get random sample of size $x \cdot |V|$ from V. Call it V_x.
3. Calculate similarity between P, P_{V_x} for $x \in (0, 1]$.
4. Repeat k times. (In this experiment, $k = 30$).
5. If the measure satisfies size monotonicity, the similarity should not decrease with sample size.

Experiment: Sample Size Monotonicity

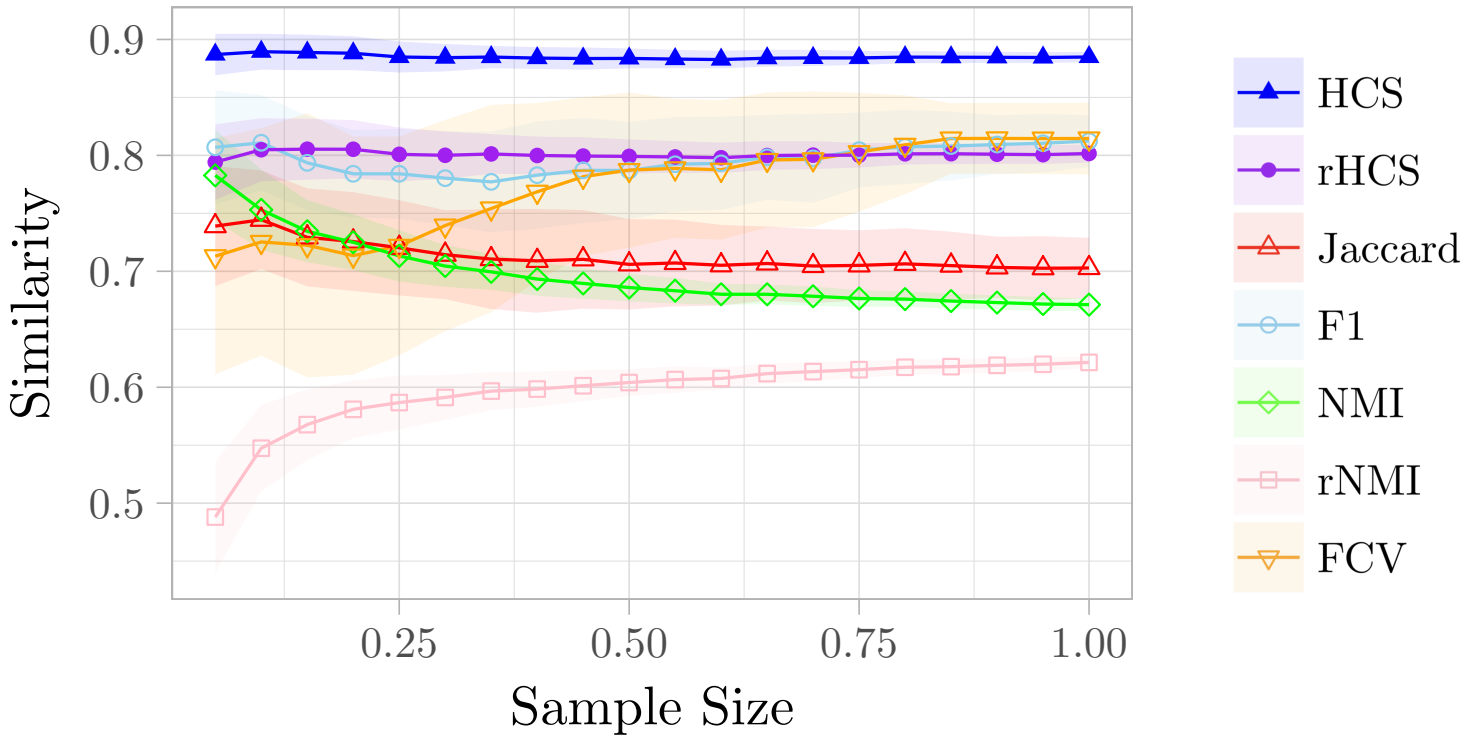


Figure: Results for Soc_Hamster

HCS and rHCS satisfies *Sample Size Independence*, and consequently *Sample Size Monotonicity*.

Among the other measures, Only NMI and Jaccard does not satisfy the *Sample Size Monotonicity*.

Thank You

Questions?

rlaishra@syr.edu