Measuring Directed Triadic Closure with Closure Coefficients

Stanford University

Hao Yin, Austin R. Benson, Johan Ugander

yinh@stanford.edu, arb@cs.cornell.edu, jugander@stanford.edu

Cornell University

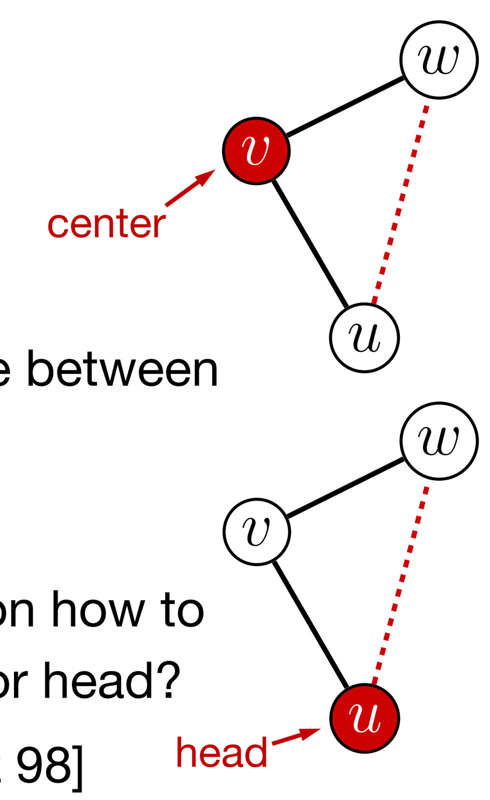
link to paper

Overview

We propose a new family of metrics for directed triadic closure, which focuses on the "initiator" of a wedge instead of the "center" node.

- Empirical analysis on the metrics exhibits an interesting block structure and a counter-intuitive asymmetry.
- Theoretical analysis under the directed configuration model partly explains the empirical observations, connecting the metrics with the degree distribution of the network.
- We demonstrate that these metrics can be powerful features in network-based machine learning tasks.

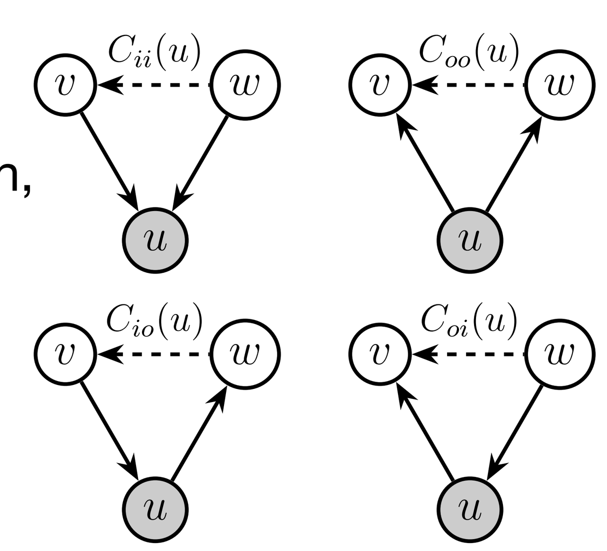
Background



❖ The clustering phenomenon

- **Observation** An increased chance of edge existence between nodes with a common neighbor (aka, triadic closure).
- **Explanation** "A friend of my friend is my friend."
- **Metrics for undirected networks** A recent debate on how to measure the rate of triadic closure: based on center or head?
  - ➢ center: the clustering coefficient [Watts & Strogatz 98]
  - ➢ head: the closure coefficients [Yin et al. 19].

❖ Directed triadic closure



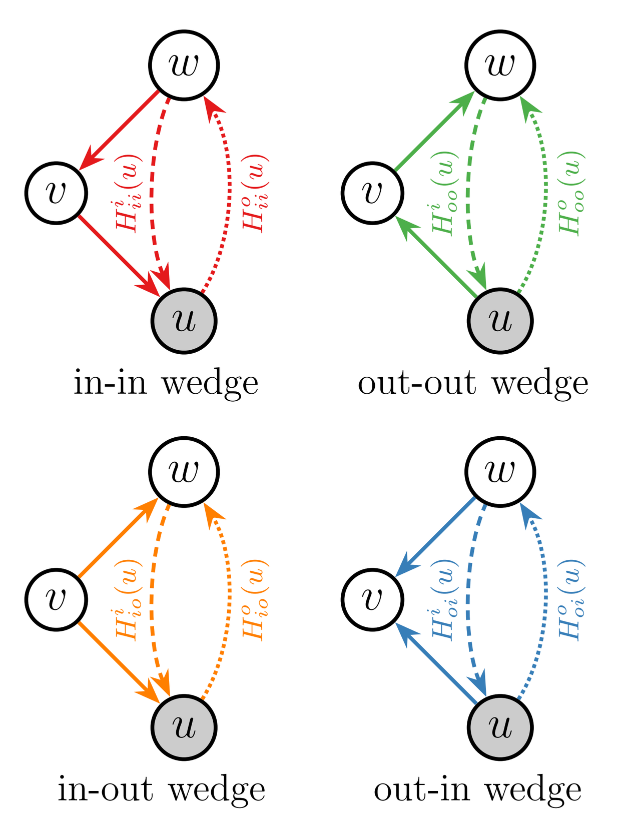
- **Idea** Edge direction is crucial in the explanation, analysis, and measurement of triadic closure.
- **Only center-based metric** Directed clustering coefficients: a four-dimensional node-level feature [Fagiolo 07].

Definition

❖ *Local* directed closure coefficients

$$H^z_{xy}(u) = T^z_{xy}(u) \,/\, W_{xy}(u).$$

- $x, y, z \in \{i, o\}$: direction of edges.
- Can be naturally generalized to consider bidirected edges separately.



❖ *Average* directed closure coefficients

$$\overline{H}^z_{xy} = \frac{1}{|V|}\sum_{u\in V} H^z_{xy}(u).$$

Empirical analysis

❖ Block structure in local directed closure coefficients

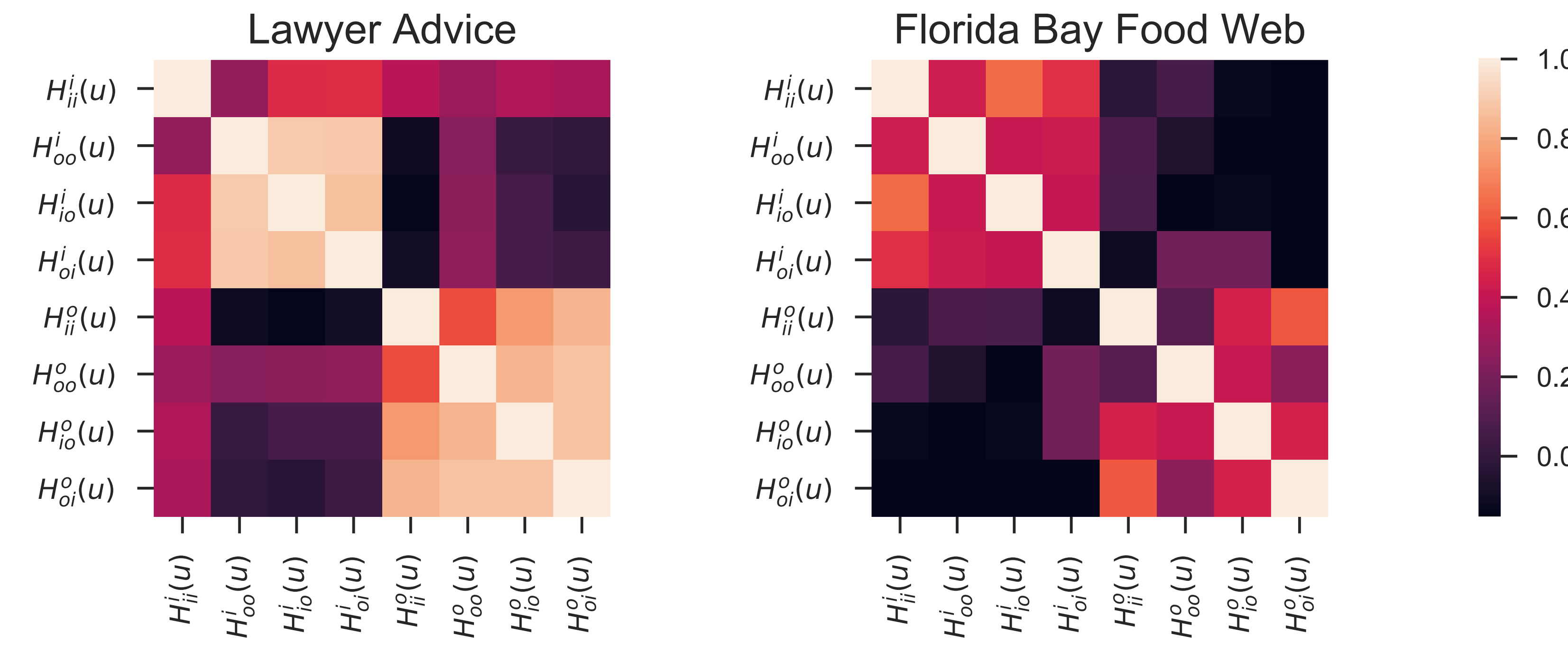
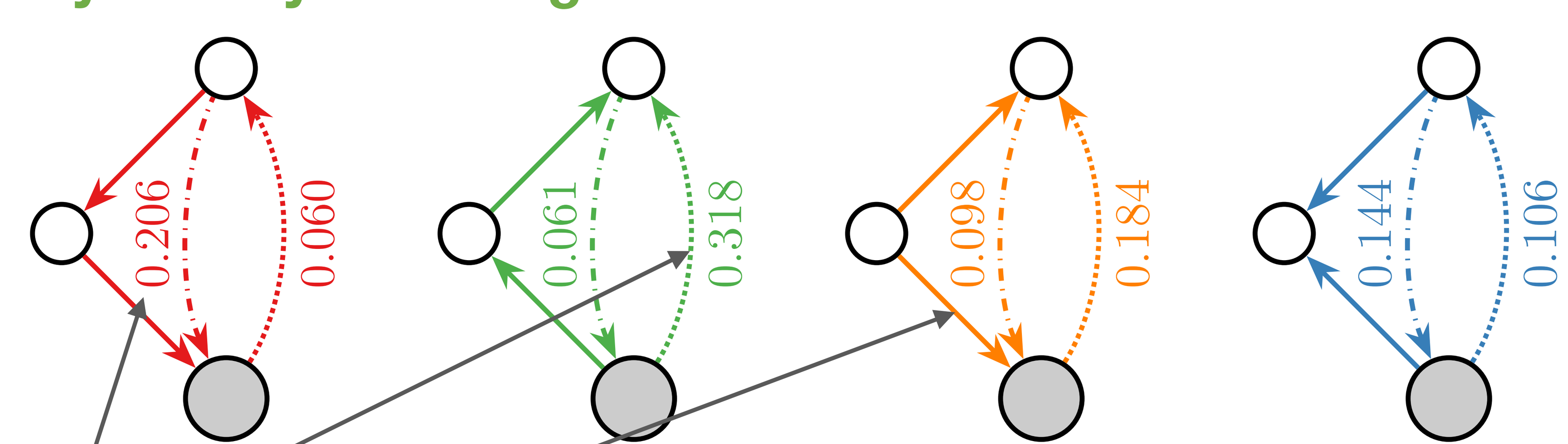


Figure: Heatmap of the correlation matrix of the eight local directed closure coefficients. We observe a clear separation on the eight closure coefficients by the inbound/outbound closure direction. Coefficients within each group are highly correlated while between groups are almost uncorrelated.

❖ Asymmetry in average directed closure coefficients



$\overline{H}^i_{ii} \ll \overline{H}^o_{oo}$, $\overline{H}^i_{io} \ll \overline{H}^o_{io}$: counter-intuitive asymmetry (Florida Bay Food Web)

Theoretical analysis

❖ Background The directed configuration model

- A uniform distribution over graphs with a specified joint in- and out-degree sequence $S = [\,(d_i(u), d_o(u))\,]_{u\in V}$.
- A degree sequence is characterized by the second-order moments $\{M_{ii}, M_{io}, M_{oo}\}$ where $M_{xy} = \frac{1}{|V|}\sum_{u\in V} d_x(u)\cdot d_y(u)$.

❖ Theorem For any joint degree sequence S and any *local* directed closure coefficient, we have

$\bar{y}, \bar{z}$: the opposite direction of y and z.

$$\mathbb{E}[H^z_{xy}(u)|S] = \frac{n(d_z(u) - \mathbf{1}_{[x=z]})}{m^2}\cdot\left(M_{\bar{y}\bar{z}} - \mathbf{1}_{[y=z]}\cdot\frac{m}{n}\right)\cdot(1+o(1)).$$

- Local directed closure coefficients positively correlates with the node degree in the closure direction!
- Local directed closure coefficients with the same closure direction should be positively correlated, which explains the block structure in the heatmap.

❖ Theorem For any joint degree sequence S and any *average* directed closure coefficient, we have

$$\mathbb{E}[\overline{H}^z_{xy}|S] = \frac{m - n\cdot\mathbf{1}_{[x=z]}}{m^2}\cdot\left(M_{\bar{y}\bar{z}} - \mathbf{1}_{[y=z]}\cdot\frac{m}{n}\right)\cdot(1+o(1)).$$

- Average closure coefficient associates with certain second-order moment of the joint degree sequence.
- Consequently, one should find no surprise for $\overline{H}^i_{io} \ll \overline{H}^o_{io}$:
  - ➢ $\overline{H}^i_{io}$ associates with $M_{io} = 201.92$;
  - ➢ $\overline{H}^o_{io}$ associates with $M_{ii} = 493.08$.

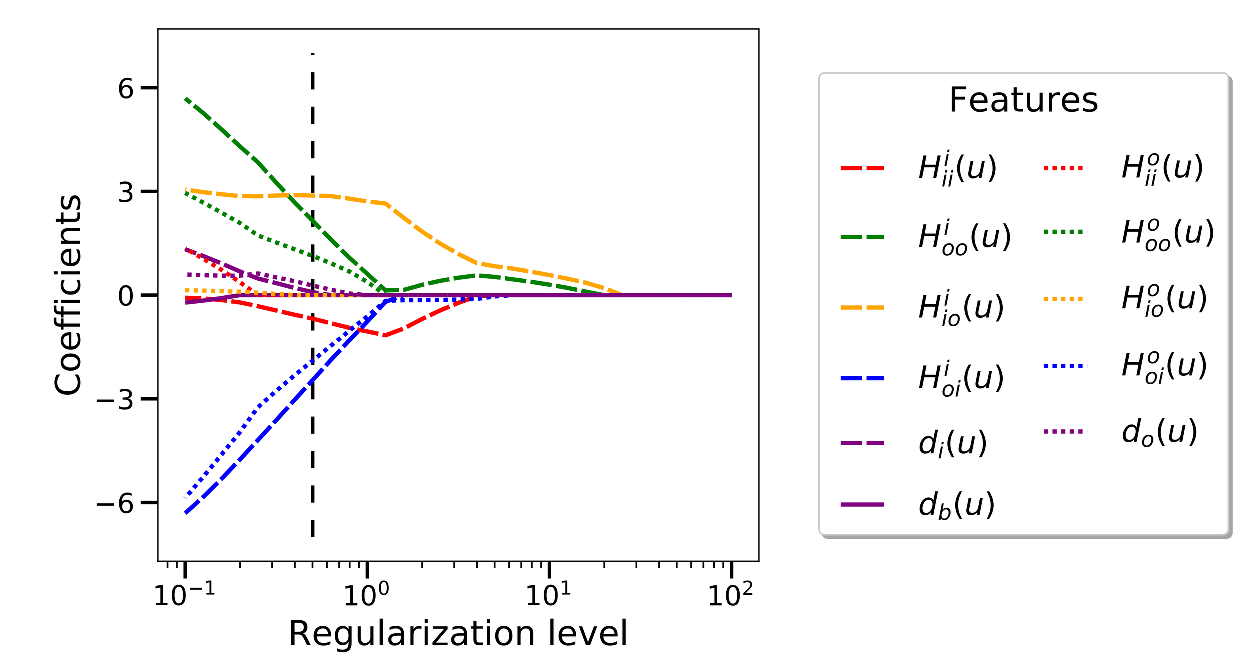
and similarly $\overline{H}^i_{ii} \ll \overline{H}^o_{oo}$.

Case study

- **Task** To predict the status of a lawyer, partner or associate?
- **Method** ℓ_1-regularized logistic regression + 3-fold CV.
- **Compare** Five feature sets:
  - Only node degrees
  - Only local directed closure coefficients
  - Combine degrees and closure coefficients
  - Only the center-based clustering coefficients [Fagiolo 07]
  - Combine degrees and clustering coefficients [Fagiolo 07]
- **Validation set accuracy and AUC**

	degree	closure	closure + degree	clustering	clustering + degree
accuracy	0.7887	**0.8723**	0.8574	0.6296	0.7886
AUC	0.8742	**0.9252**	0.9166	0.6446	0.8746

- **Model interpretation**



Key References

1. Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." *Nature* 393.6684 (1998): 440.
2. Yin, Hao, Austin R. Benson, and Jure Leskovec. "The Local Closure Coefficient: A New Perspective On Network Clustering." *WSDM* 2019.
3. Fagiolo, Giorgio. "Clustering in complex directed networks." *PRE* 76.2 (2007): 026107.