Growing Attributed Networks Through Local Processes

Harshay Shah, Suhansanu Kumar, and Hari Sundaram
{hrshah4, skumar56, hs1}@illinois.edu

THE WEB CONFERENCE

ILLINOIS

The Problem

Existing network growth models often make assumptions that are at variance with how individuals form links in real-world networks:

1. Individuals utilize unbounded computation and information (e.g., node degree) to form links
2. Individuals only rely on structural features to form links and do not consider nodal attributes
3. Individuals form each link independently

We address the following problem:

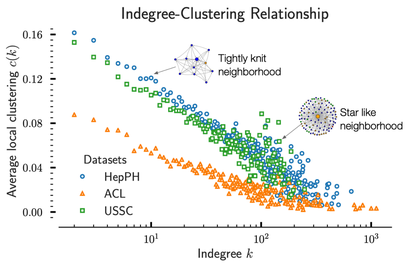
> How do individual resource constraints shape the global structure of attributed networks?

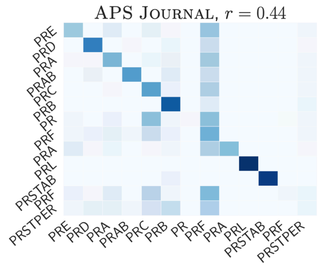
Empirical Analysis

We study evolving attributed networks in which new nodes, each with an attribute value, join the network and form edges to existing nodes.

Global Network Properties

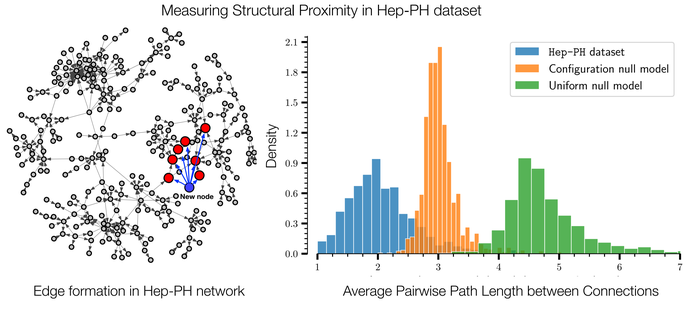
Common trends in six large-scale citation networks: densification, high local clustering, heavy-tailed degree distribution, and homophily.





Locality-biased Edge Formation

Edge formation processes exhibit bias towards pairwise proximate nodes that are in the same locality.



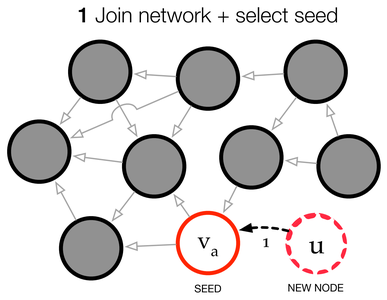
Our Contributions

We propose a network growth model that explains how key structural properties of attributed networks can jointly emerge from a local edge formation process. Our model is normative, accurate and interpretable:

- **Normative** — incorporate sociological phenomena that describe how individuals form edges under resource constraints
- **Accurate** — preserve key structural properties and attribute mixing patterns of real-world attributed networks
- **Interpretable** — generate attributed networks with tunable structural properties using four, easy-to-understand parameters

Attributed Random Walk (ARW) Model

ARW is a resource-constrained model in which new nodes use biased random walks to concurrently acquire local information and form edges.

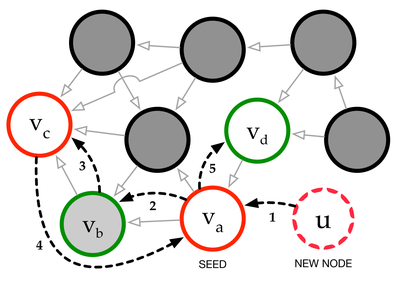


Select-Seed

(1) With probability $p_{\text{same}}/p_{\text{same}}+p_{\text{diff}}$, randomly select a seed node from existing nodes that have the same attribute value, $B(u)$.

(2) Otherwise, with probability $p_{\text{diff}}/p_{\text{same}}+p_{\text{diff}}$, randomly select a seed node from existing nodes that do *not* have the same attribute value, $B(u)$.

ARW parameters jointly control random walk traversal, link formation, and seed selection.



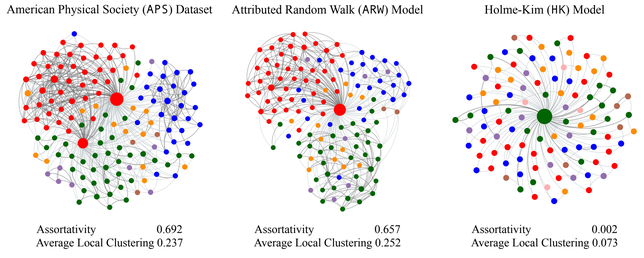
Random-Walk

(1) At each step of the walk, new node u visits node v_i.
- If $B(u) = B(v_i)$, u links to v_i with probability p_{same}
- Otherwise, u links to v_i with probability p_{diff}

(2) Then, with probability p_{jump}, u jumps back to seed s_u.

(3) Otherwise, with probability $1 - p_{\text{jump}}$, u continues to walk. It picks an outgoing edge with prob. p_{out} *or* an incoming edge with prob. $1 - p_{\text{out}}$ to visit a neighbor of v_i.

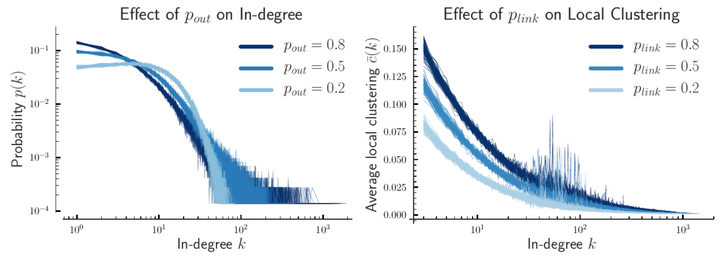
(4) Steps 1-3 are repeated until u links to $m(t)$ nodes.



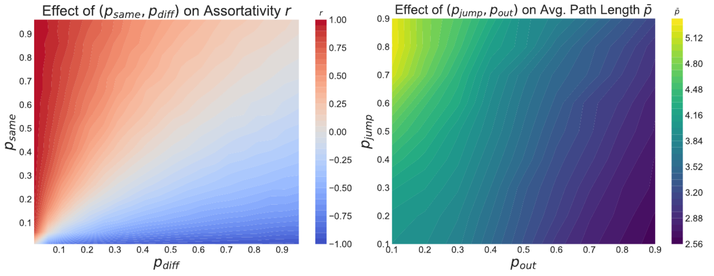
Contrasting ARW with the Holme-Kim model underscores the importance of incorporating multiple sociological phenomena in edge formation processes.

Interpreting the Model

The model parameters control the probability of connecting to similar, proximate & high-degree nodes and subsequently shape structural properties.

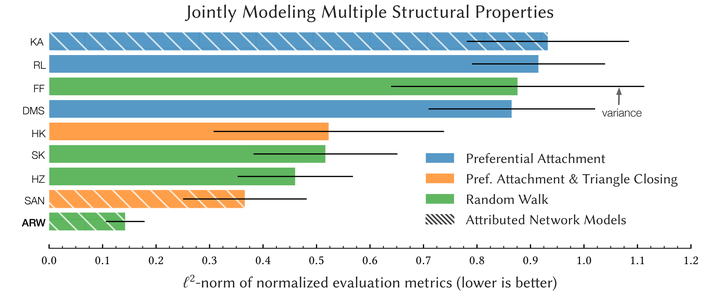


Note that the link formation parameters tune attribute mixing patterns as well as local clustering.



Evaluation

ARW outperforms eight well-known models in jointly preserving global network properties—attribute assortativity, degree, local clustering—by a statistically significant margin of 2.5-10x.



Future Directions

- Resource-constrained models for social networks
- Motif formation in attributed networks
- Edge formation in the presence of multiple attributes

Acknowledgements This project was supported in part by the Office of Undergraduate Research (OUR) at UIUC.