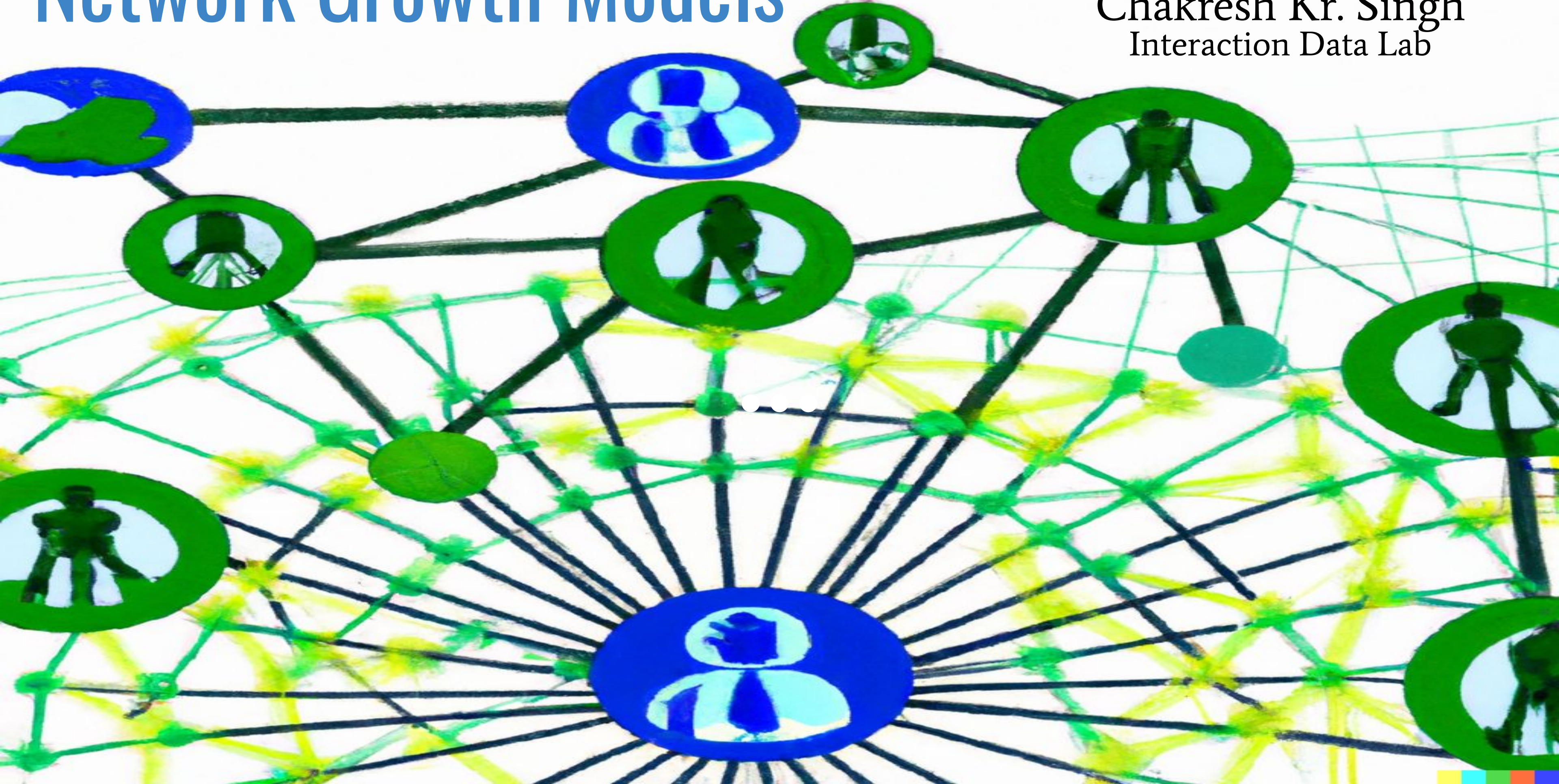
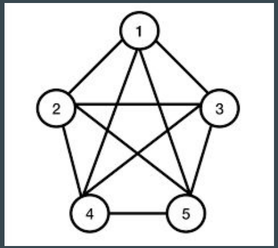
Network Growth Models

Chakresh Kr. Singh
Interaction Data Lab

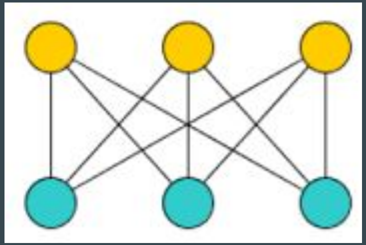


Different Network Types

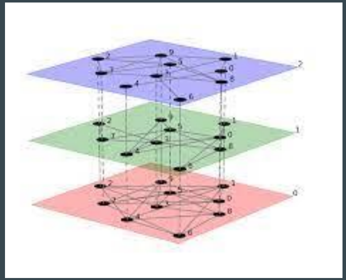
- Unipartite



- Bipartite

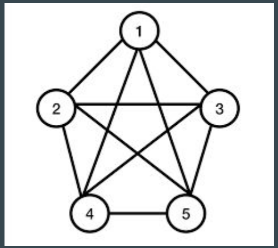


- Multilayer

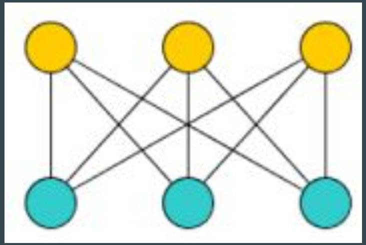


Different Network Types

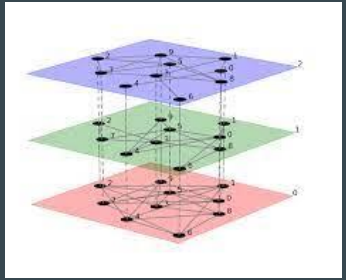
- Unipartite



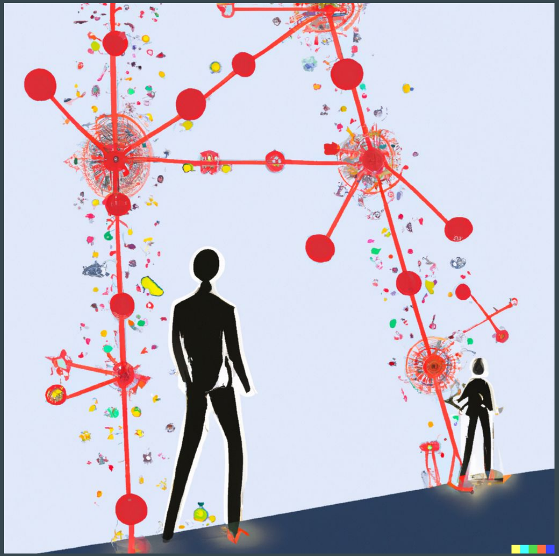
- Bipartite



- Multilayer



On the shoulder of Giants



Mathew Effect
Simon's Model
Scalefree
Preferential
Polya's Urn Model
Power Law Barabasi-Albert

2000
1955
Yule-process 1924
Zipfs Law
Evolution

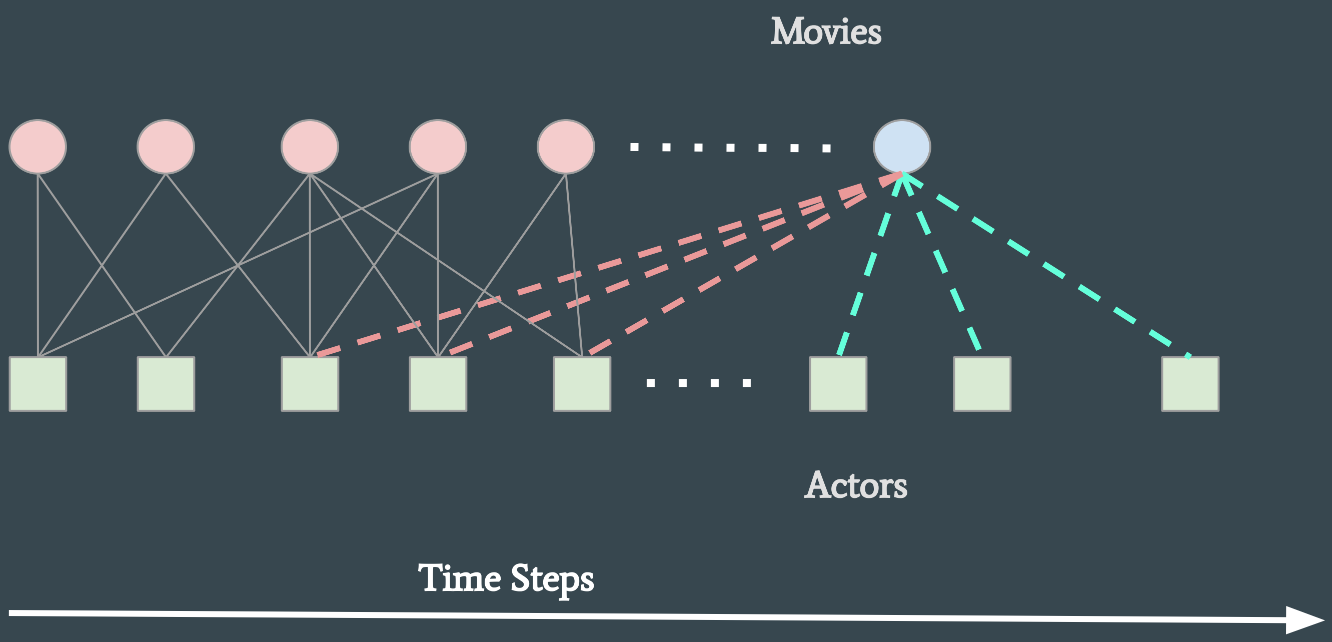
Our Interests

- IGEM - members to wiki edits
- arXiv - authors to Scientific Papers
- GitHub - user to Files

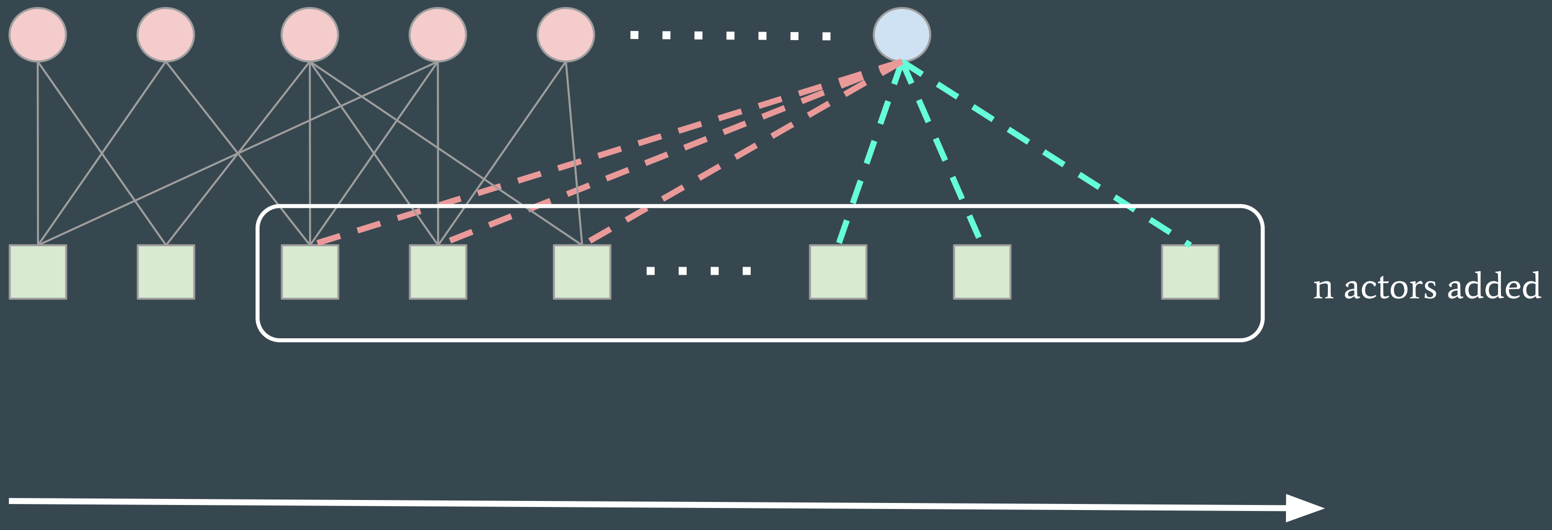
Different Approaches to network models

- Generative Models
    - When we define steps of nodes/edges addition based on some probability distributions
- Ecological Models
    - A lot of them are based on Master equation defining interaction between
- Machine Learning
    - Prediction of new links

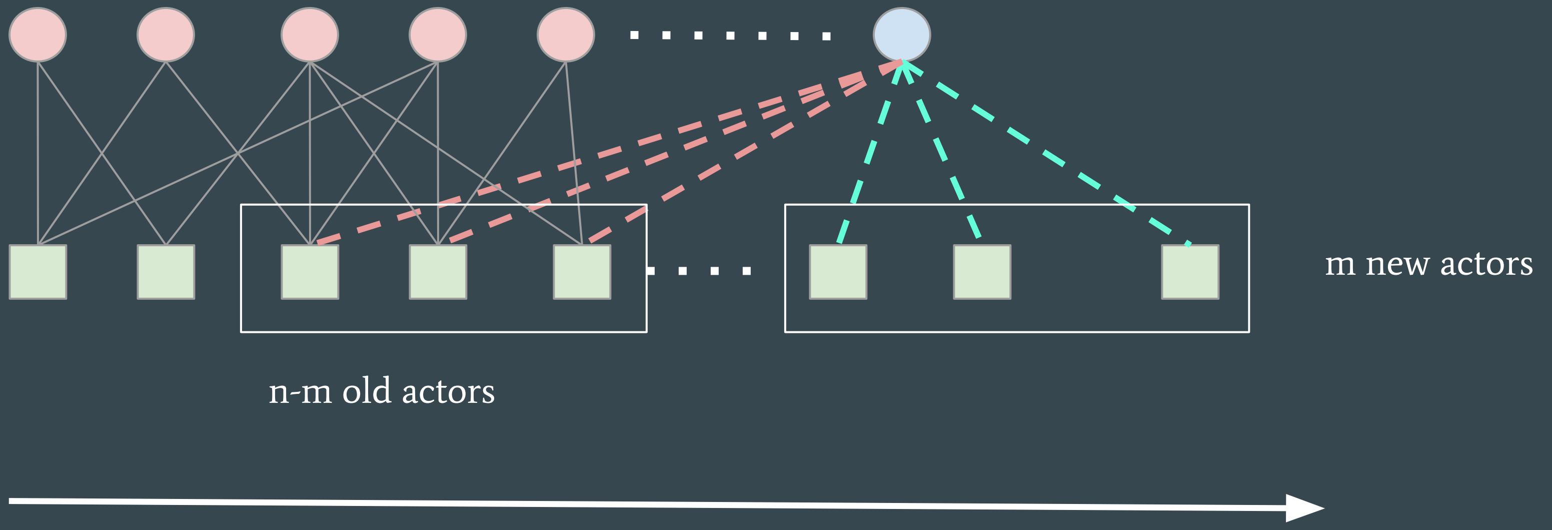
Ramasco’s Model (Self organization in Collaboration Networks)



Ramasco's Model (Self organization in Collaboration Networks)



Ramasco's Model (Self organization in Collaboration Networks)

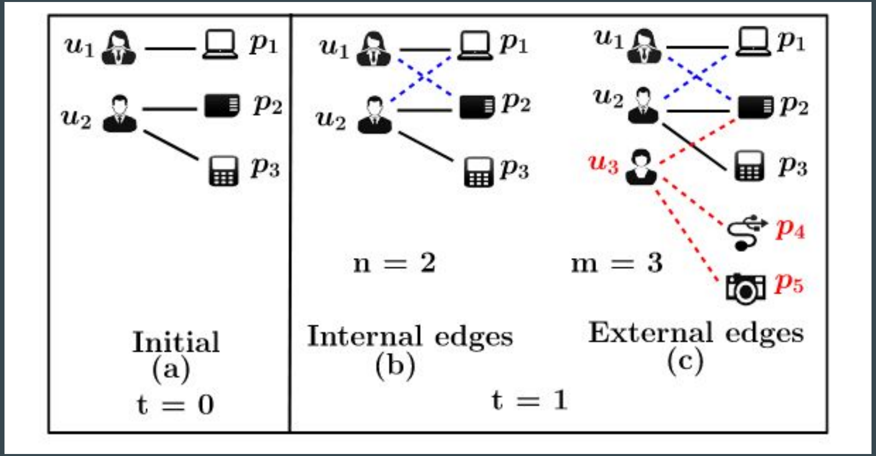


Playing around with Bipartite - Models

Physica A: Statistical Mechanics and its Applications
Volume 517, 1 March 2019, Pages 370-384

A general growth model for online emerging user–object bipartite networks ☆

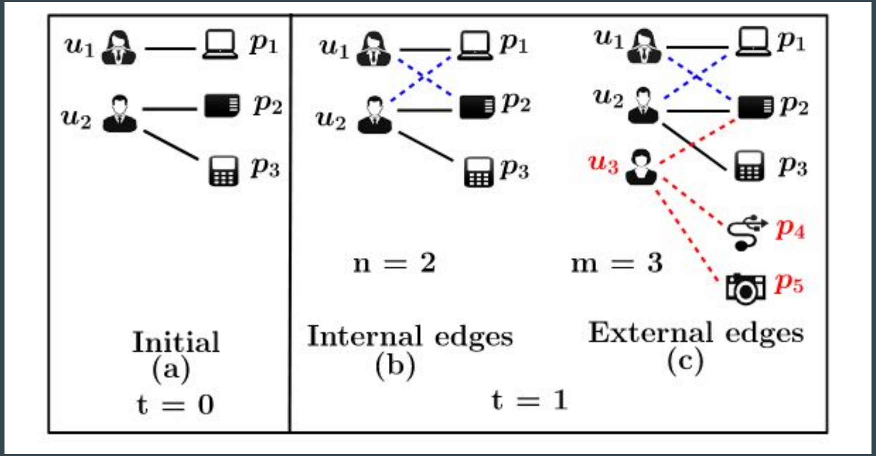
Anita Chandra, Himanshu Garg, Abyayananda Maiti



Growth of the Model

$$\widetilde{A}(k_{v,t}) = \frac{k_{v,t} + \gamma}{\sum_{v=1}^{N}(k_{v,t} + \gamma)}$$

Attachment Kernel



Growth of the Model

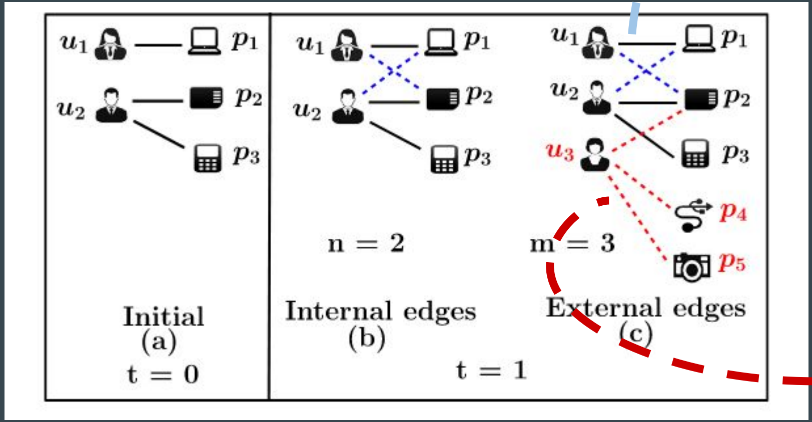
- We can test other attachment kernels
- Custom functions?

$$\widetilde{A}(k_{v,t}) = \frac{k_{v,t} + \gamma}{\sum_{v=1}^{N} (k_{v,t} + \gamma)}$$

Attachment Kernel

Probability of connection for internal edges

$$\frac{k_v + \gamma_i}{\sum_{v=1}^{o_0+wt}(k_v + \gamma_i)}.$$



Probability of connection for external edges

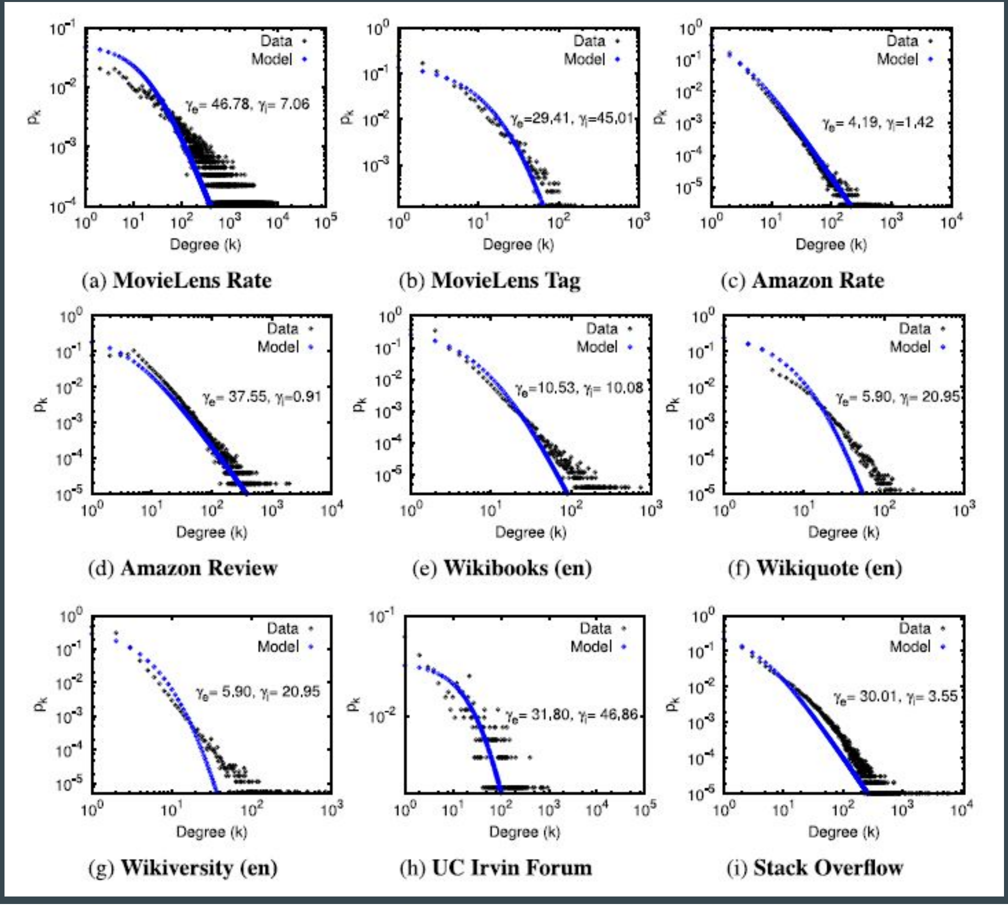
$$\frac{k_v + \gamma_e}{\sum_{v=1}^{o_0+wt}(k_v + \gamma_e)}.$$

This is a little different from Ramasco's Model (movie-actor)

$$\frac{\partial k_v}{\partial t} = m\frac{k_v + \gamma_e}{\sum_{v=1}^{o_0+wt} k_v + \gamma_e} + n\frac{k_v + \gamma_i}{\sum_{v=1}^{o_0+wt} k_v + \gamma_i}.$$

$$\frac{\partial k_v}{\partial t} = m\frac{k_v + \gamma_e}{(c + w\gamma_e)t} + n\frac{k_v + \gamma_i}{(c + w\gamma_i)t}.$$

c = m+n

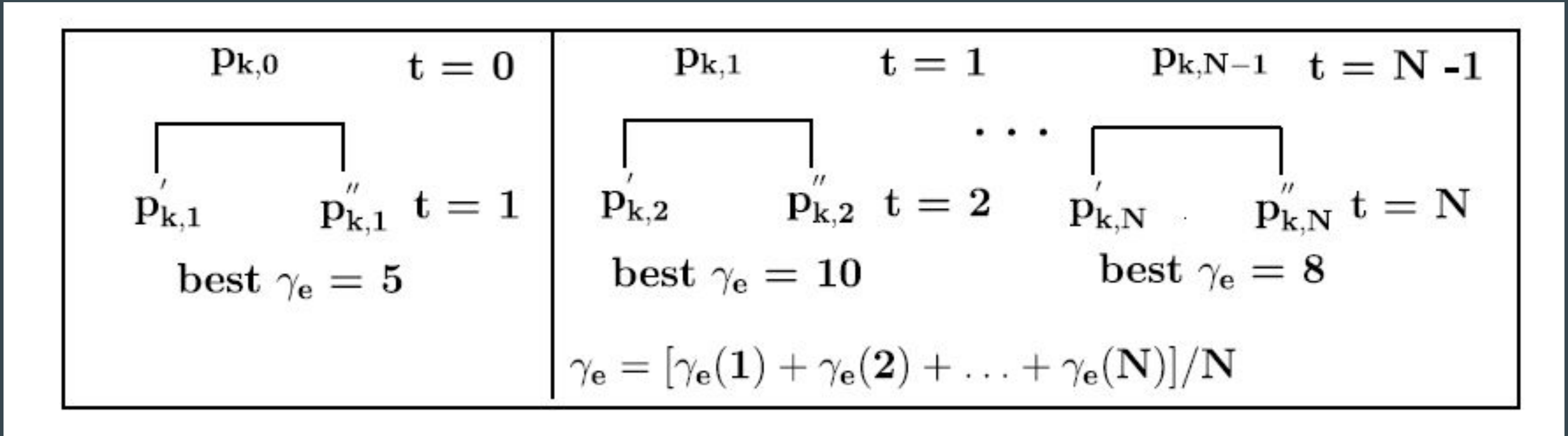


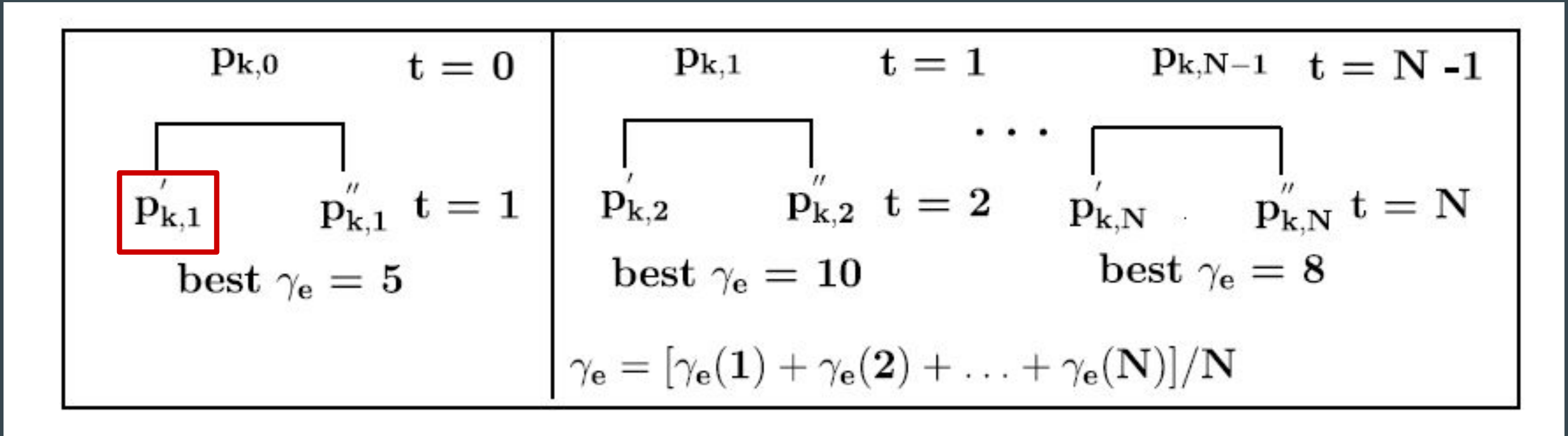
$$p(k) = \frac{\partial P(k_v(t) < k)}{\partial k} = \frac{r}{s}(k_0)^{\frac{r}{s}}(k + k_0)^{-(1+\frac{r}{s})}.$$

They end up with a shifted power law

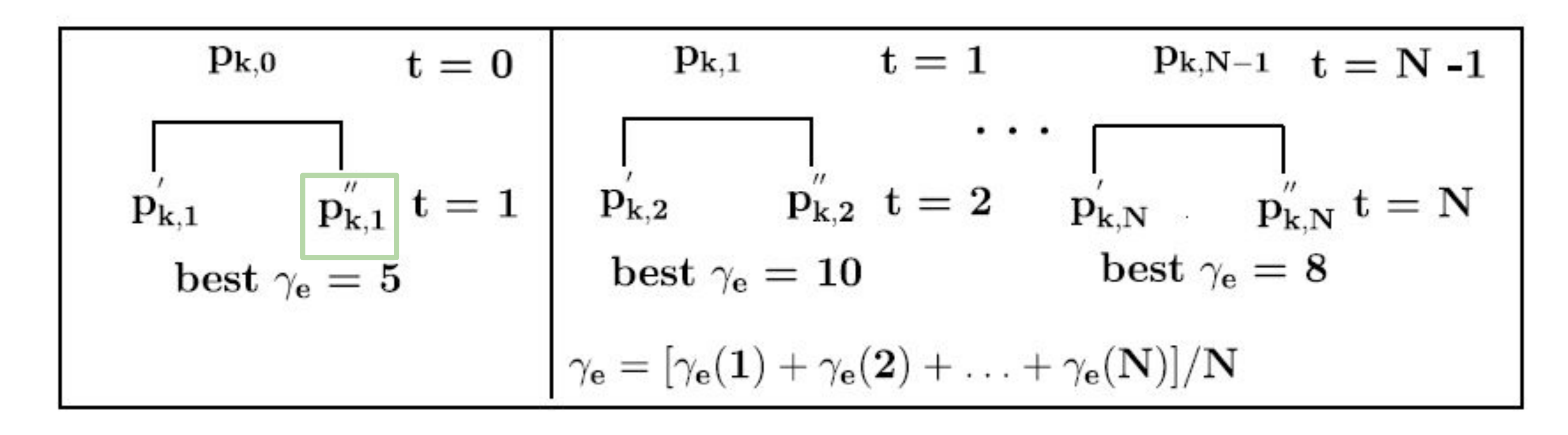
Fig. 5. Object degree distribution plots for nine different online networks from real data and model for (a) movies of MovieLens (rate), (b) movies of MovieLens (tag), (c) products of Amazon (rate), (d) electronic products of Amazon (review), (e) books of Wikibooks (en), (f) quotes of Wikiquote (en), (g) learning projects or resources of Wikiversity (en), (h) forums of UC Irvin Forum, and (i) Q & A posts of Stack Overflow. X-axis shows the degree (k) and Y-axis depicts the probability of having an object with degree k (p_k).

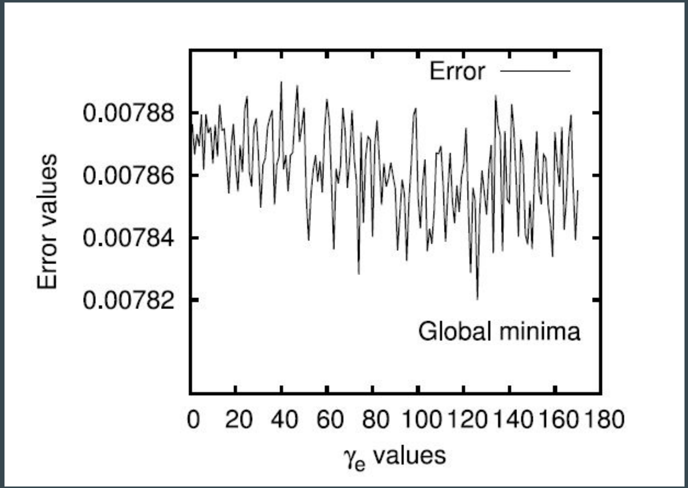
Estimating the randomness for every user





p'(k,t+1) - derived from original p(k,t) after excluding n internal edges arriving at t+1





- p”(k,t+1) - derived from original p(k,t) by simulating only considering the m edges.
- Then the difference b/w p’(k,t+1) and p”(k,t+t) is minimized to estimate Ye

Similarly considering only n internal edges they estimate Yi

Take Away

- The model is flexible to different attachment kernels
- The estimation of randomness parameter ye and yi differentiate between user choice of preferential vs random behavior... also telling which user is what.

To Consider

- The appearance of n edges is not clearly defined and could have been better
- It would have been better to use KL-div test statistic to compare distributions

Link Prediction

EPJ.org

REGULAR ARTICLE Open Access

Check for updates

Prediction of new scientific collaborations through multiplex networks

Marta Tuninetti[1†], Alberto Aleta[1†], Daniela Paolotti[1], Yamir Moreno[1,2,3] and Michele Starnini[1*]

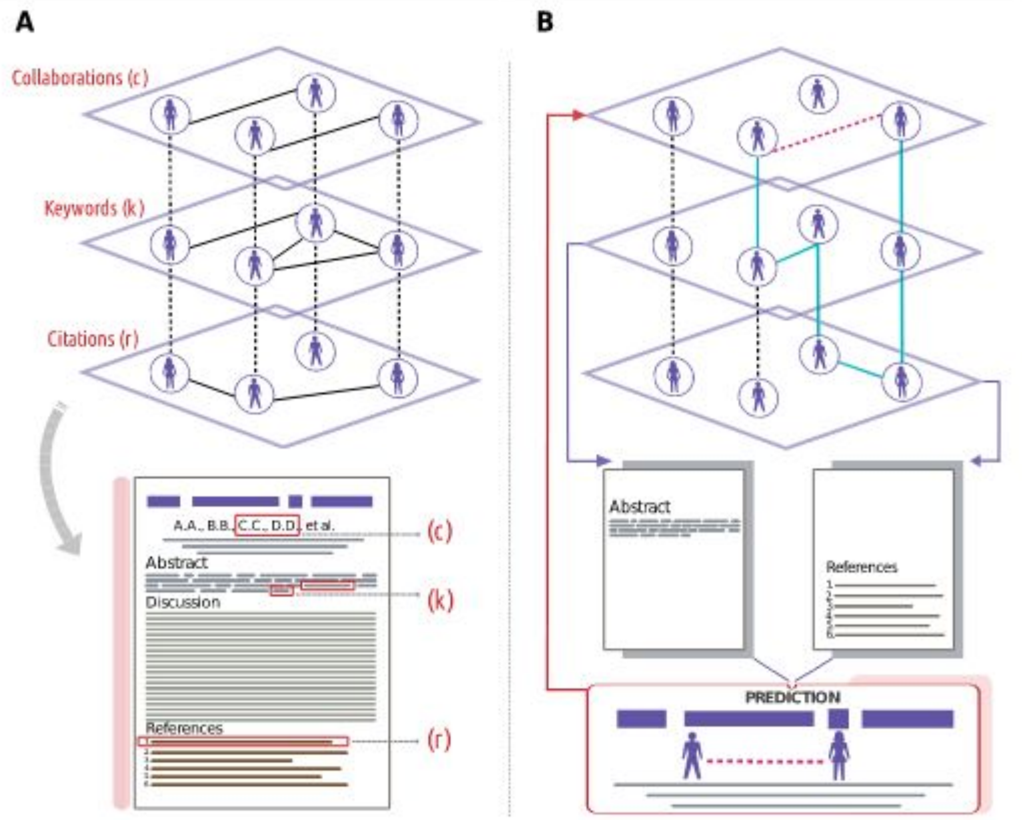


Figure 1 *Prediction of new collaborations using multiplex networks.* We first build a multiplex network using three different kinds of relational data between scientists (panel **A**) and then use two of them (panel **B**) to predict new collaborations among them. See the text for further details

Exploiting Diadic relationships

$$MC(u,v) = \boxed{w^r_{uv}.}$$

Mutual citations

weight of link b/w u,v in the Citations layer

$$CK(u,v) = \boxed{w^k_{uv}.}$$

Common keywords

weight of link b/w u,v in the Keywords layer

Exploiting Diadic relationships

$$NMC(u, v) = \frac{w^r_{uv}}{s^r_u} + \frac{w^r_{vu}}{s^r_v},$$

Normalized

Mutual citations $$MC(u, v) = w^r_{uv}.$$

Common keywords $$CK(u, v) = w^k_{uv}.$$

Normalized

$$NCK(u, v) = \frac{w^k_{uv}}{\max(K_u, K_v)},$$

Triadic Relationship

$$AA_\alpha(u,v) = \sum_{w \in \Gamma_\alpha(u) \cap \Gamma_\alpha(v)} \frac{1}{\ln(k_w^\alpha)},$$

Adamic Ader (Paper)

w is the set of common nbrs of u and v

k_w is the degree of w in layer alpha

Triadic Relationship

$$AA_\alpha(u,v) = \sum_{w \in \Gamma_\alpha(u) \cap \Gamma_\alpha(v)} \frac{1}{\ln(k_w^\alpha)},$$

Adamic Ader (Paper)

w is the set of common nbrs of u and v

k_w is the degree of w in layer alpha

$$MAA(u,v) = \sum_{\alpha,\beta} \sum_{w \in T_{\alpha\beta}} \frac{\eta_{c\alpha}\eta_{c\beta}}{\sqrt{\langle k \rangle_\alpha \langle k \rangle_\beta}} \frac{1}{\sqrt{\ln(k_w^\alpha)\ln(k_w^\beta)}},$$

Adamic Ader like measure for multiplex networks (MAA) (Paper)

Eta controls the respective weight of each layer with

$$\sum_\alpha \eta_{c\alpha} = \eta_{cc} + \eta_{cr} + \eta_{ck} = 1.$$

c = collaboration; r = citation; k = keywords

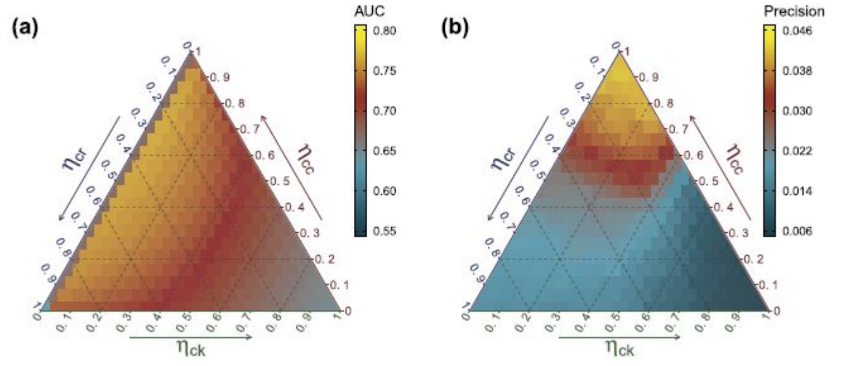


Figure 2 *AUC and Precision values of the MAA metric for different values of the coefficients* $\eta_{c\alpha}$. Varying the values of η_{cr} and η_{ck}, the third parameter η_{cc} is naturally fixed

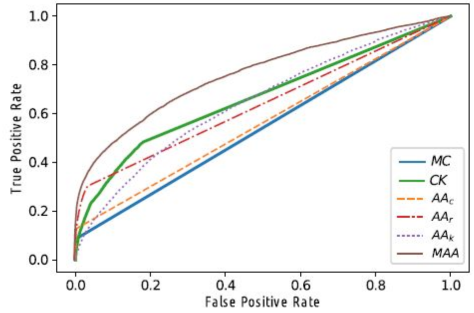


Figure 3 *ROC curves corresponding to dyadic metrics, AA single-layered scores, and the MAA metric.* In solid lines, *MC*, *CK* and *MAA* scores. The *MAA* metric obtained with coefficients $\eta_{ck} = 0.05$ and $\eta_{cr} = 0.1$ shows the best performance of the whole set of metrics. The single-layer versions of the AA metric are shown in dashed (collaboration layer), dot-dashed (citation layer) and dotted (keywords layer) lines. The normalized metrics (*NMC* and *NCK*) completely overlap their non-normalized counterparts, and thus are removed for clarity. Similarly, the *AA* metric applied over the aggregated network (not shown) completely overlaps with the AA_k curve, showing that the main contribution in the aggregated network comes from the keywords layer

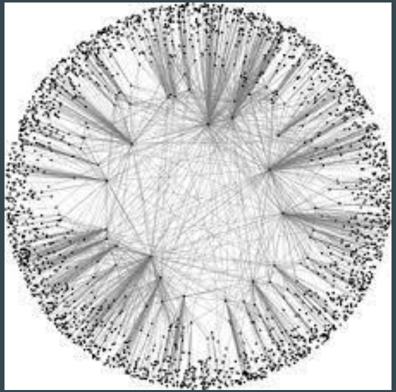
Take Away

- Easy to implement
- Extendible to multiple layers

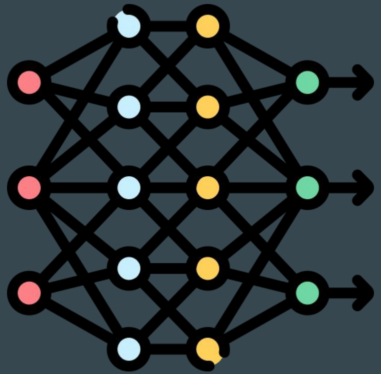
To Consider

- Very large graphs would be computationally expensive
- Redundancy in layers

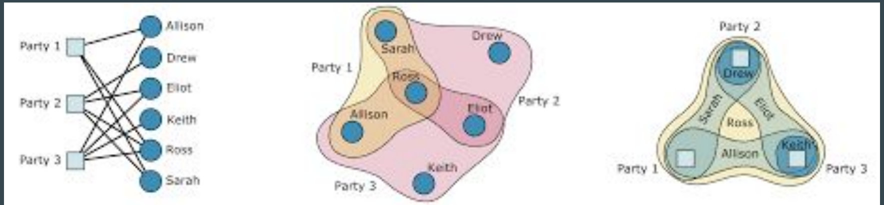
Various other popular approaches



Graph encoding in GNNs



Bipartite Growth in Hyperbolic Space



Exploiting the dynamics of higher order interactions