



Introduction to Link Prediction

Machine Learning and Modelling for Social Networks

Lloyd Sanders, Olivia Woolley, Iza Moize, Nino Antulov-Fantulin

D-GESS: Computational Social Science

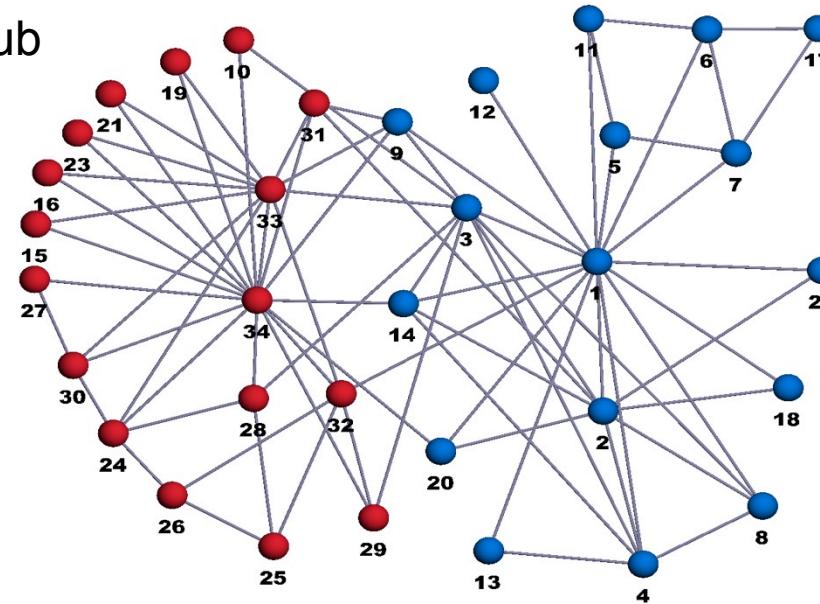
Overview

- What is link prediction?
- What are some examples of link prediction?
- A How To
 - The setup
 - Local Methods
 - Global Methods
- Other Methods
- Challenges of Link Prediction
- References

Statement of the Problem

Abstractly: Given a snapshot of a network, can one predict the next most likely links to form in the network?

Zachary Karate Club



Link Prediction in Industry (Recommender Systems)



Proposing friendships

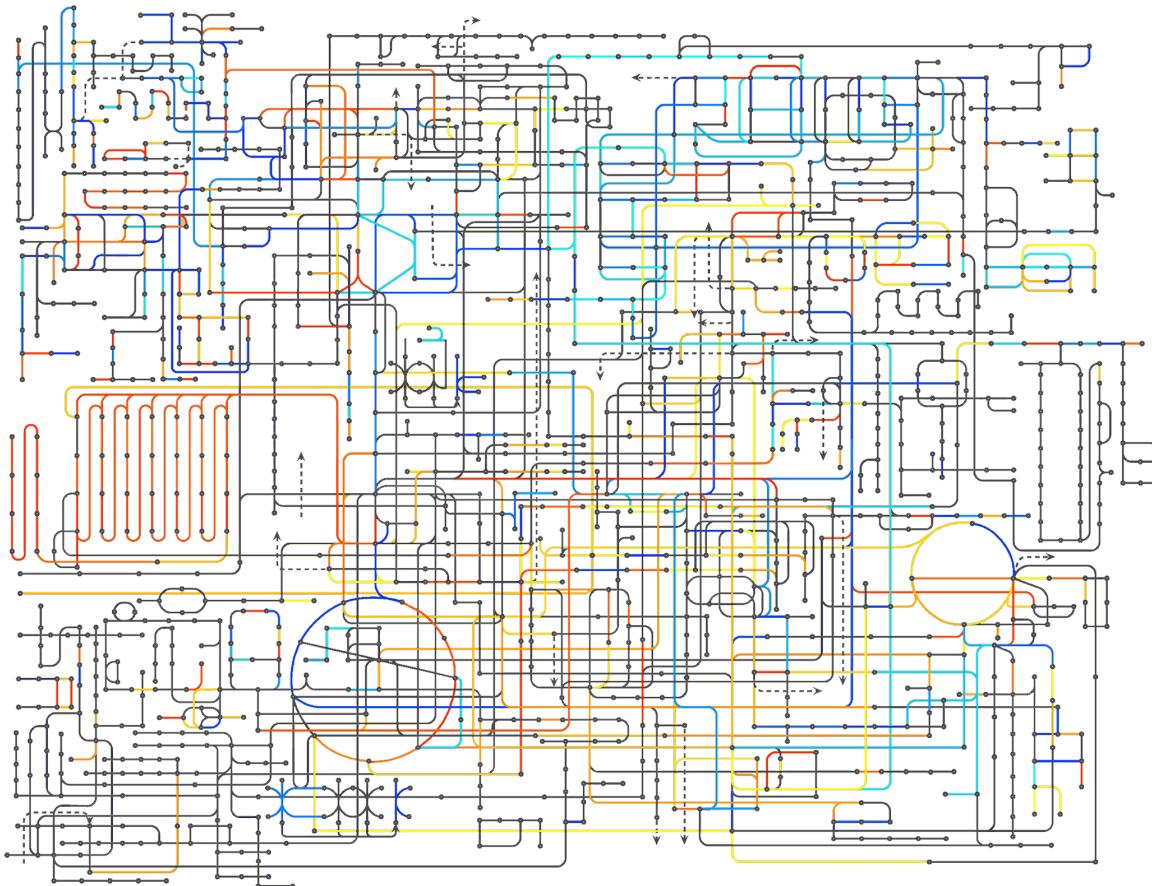


Proposing matches



Proposing items to purchase

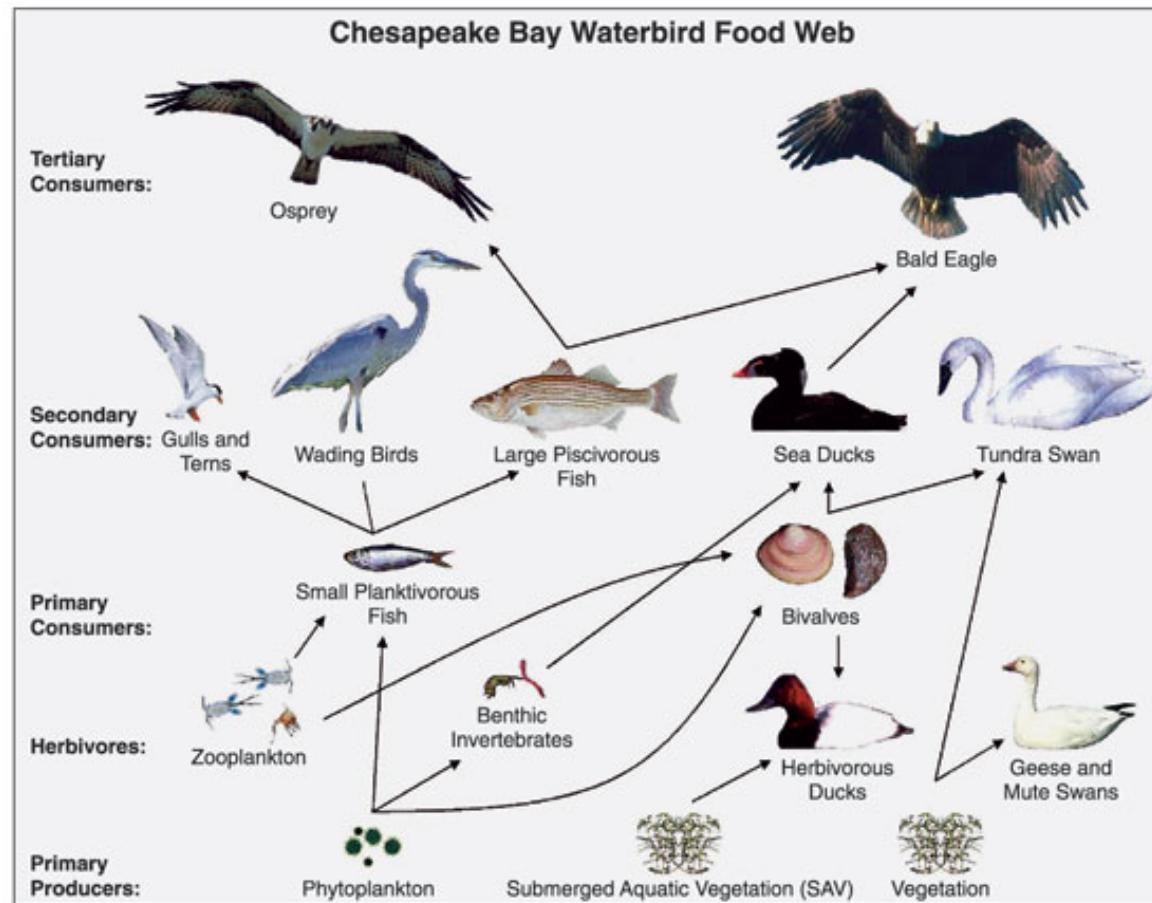
Link Prediction in Science



Artist's rendition of Human Metabolic network

Wikipedia

Link Prediction in Science



Wikipedia

Investigating link connections in bio. networks is costly and time consuming.

Link Prediction



RESEARCH ARTICLE

Link Prediction in Criminal Networks: A Tool for Criminal Intelligence Analysis

Giulia Berlusconi¹, Francesco Calderoni^{1*}, Nicola Parolini², Marco Verani², Carlo Piccardi^{3*}

1 Università Cattolica del Sacro Cuore and Transcrime, Milano, Italy, **2** MOX, Department of Mathematics, Politecnico di Milano, Milano, Italy, **3** Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy

* francesco.calderoni@unicatt.it (FC); carlo.piccardi@polimi.it (CP)

Possible missing links predicted between suspects in organized crime

Link Prediction

Link prediction is used to predict future possible links in the network (E.g., Facebook). Or, it can be used to predict missing links due to incomplete data (E.g., Food-webs – this is related to sampling that Olivia spoke of earlier).

Statement of the Problem

Abstractly: Given a snapshot of a network, can one predict the next most likely links to form in the network?

Graph

$$G = (V, E)$$

Edge

$$e = (x, y)$$

Similarity Score

$$s_{xy}$$

How To: The Setup

- For a given graph
- Split the data into a training set, and a test set
- Choose a link prediction algorithm
- Run the algorithm on the training set, and test it on the test set.
- Check the accuracy
- Compare other link prediction algorithms

How To: The Setup

Split Edges into a training and test (probe) set

$$E = E_{train} \cup E_{test}$$

Set of all possible edges on graph

$$|U| = \frac{|V|(|V| - 1)}{2}$$

How To: Output of algorithm

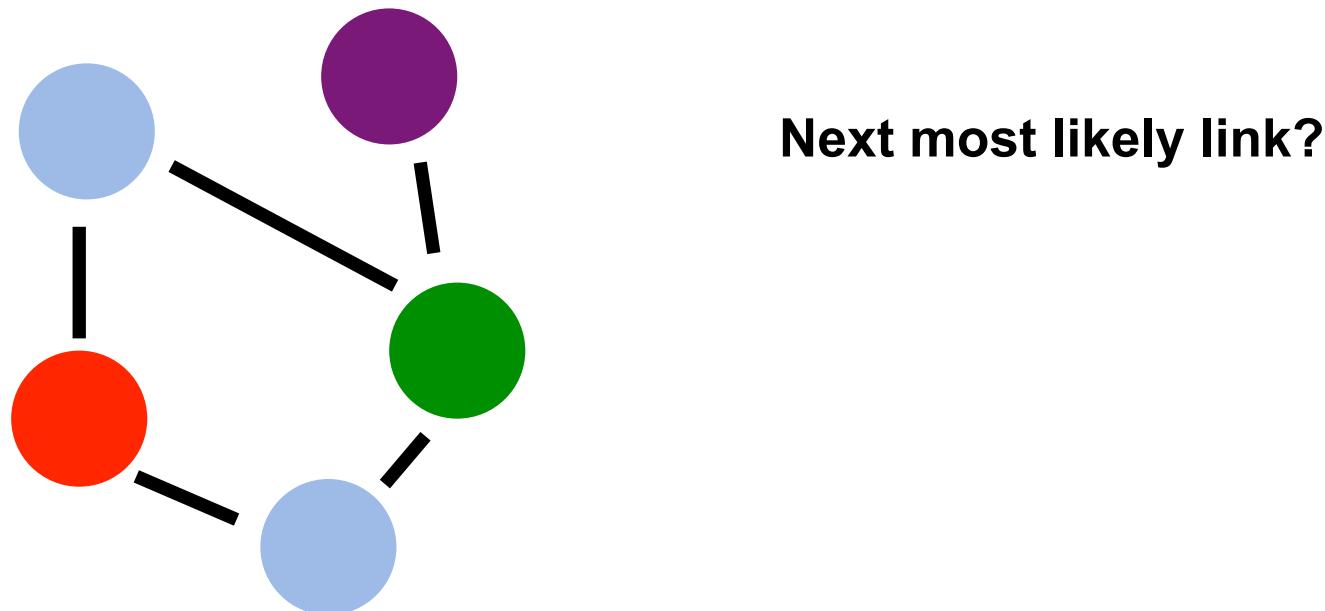
The Link Prediction algorithm will spit out a list, ranked with edges which are most likely to appear at the top, descending.

$$L : e_L \in U - E_{train}$$

Taking the first n links from the list, and calculating the intersection with the probe set of length n, gives a simple measure of accuracy

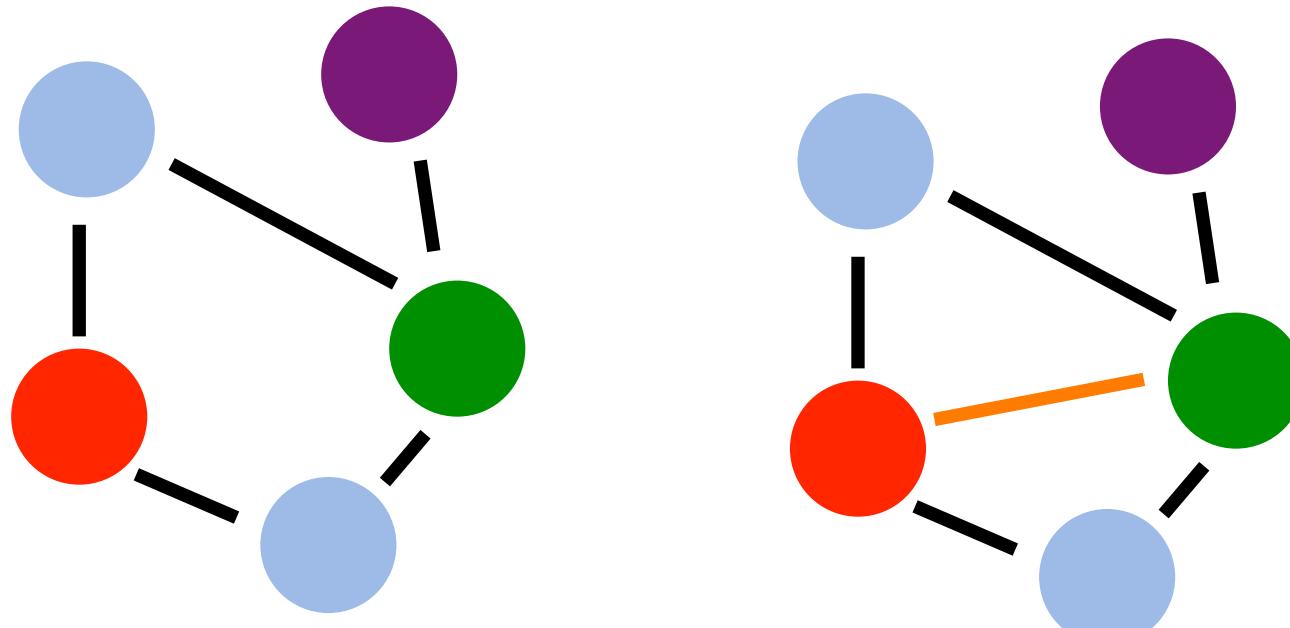
Similarity measures: Local

Local methods predict links based solely on your local contact structure. The main idea is that of triangle closing



Similarity measures: Local

Local methods predict links based solely on your local contact structure. The main idea is that of triangle closing

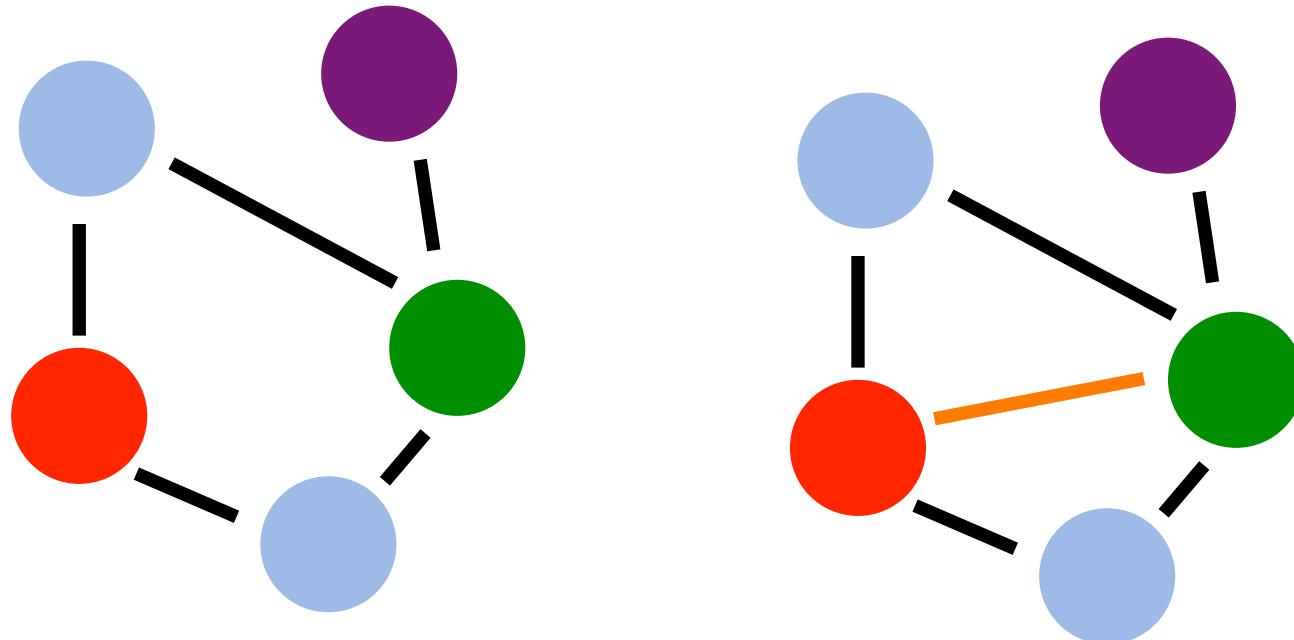


How to: Similarity Measures

Common Neighbours

$$s_{xy}^{\text{CN}} = |\Gamma(x) \cap \Gamma(y)|$$

Set of NN. Cardinality is the degree of the node



Counting number of paths of a certain length

How to: Similarity Measures

Jaccard Index

$$S_{xy}^{\text{Jaccard}} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

Resource Allocation

$$S_{xy}^{\text{RA}} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z}$$

Intuition: If we share neighbors who have a low degree, we are likely to be connected. Being connected through a hub is 'meaningless'

Many more similarity measures available

Similarity measures: Global

- Given the adjacency matrix, one can take the global structure of the graph into account when making predictions.
- These are often based on number of shortest path measures.
- Common neighbors is simply the Adj. matrix squared.

$$s_{xy}^{CN} = (A^2)_{xy}$$

Try it for yourself to verify it!

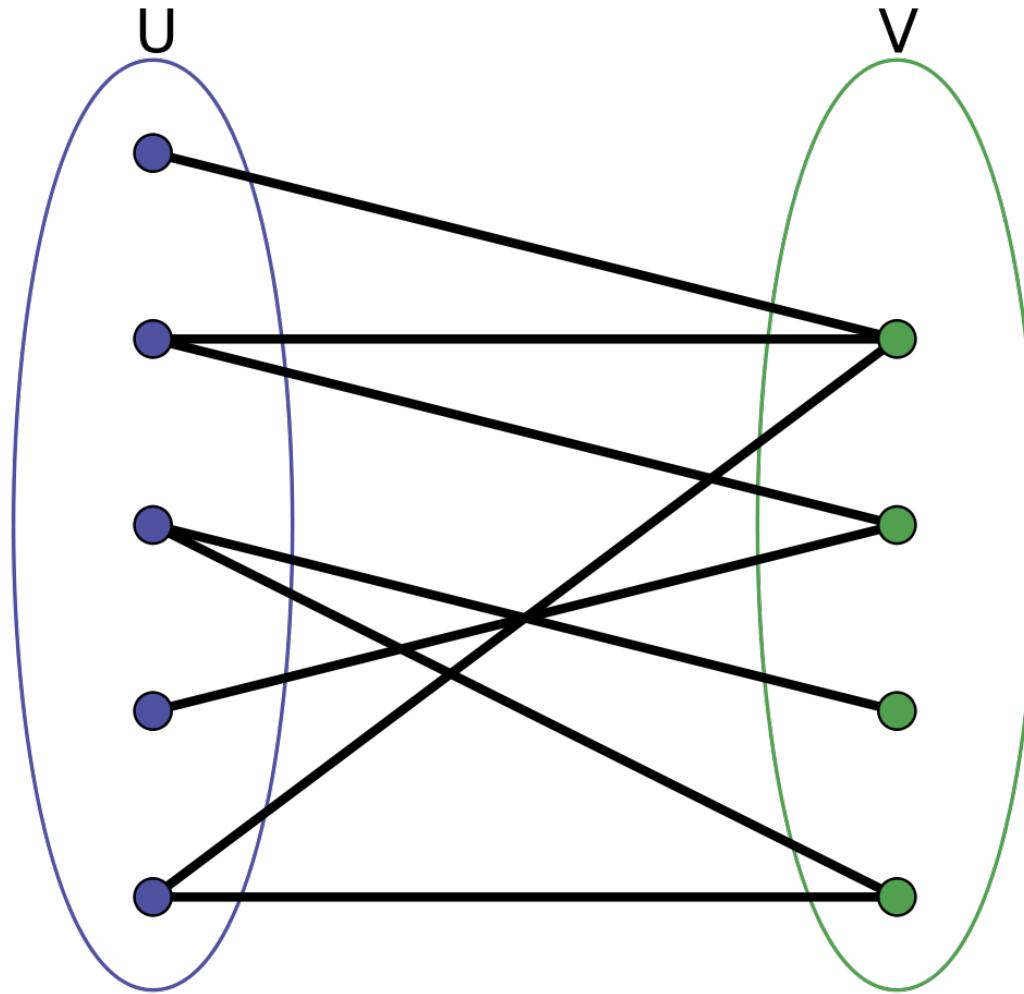
Global Methods

- What about paths of greater length? How can we include those? We can weight longer paths less.

$$\exp(\alpha A) = 1 + \alpha A + \frac{(\alpha^2 A^2)}{2!} + \frac{(\alpha^3 A^3)}{3!} + \dots$$

$$s_{xy} = \exp(\alpha A)|_{xy} = \sum_{i=0}^{\infty} \frac{\alpha^i}{i!} A^i|_{xy}$$

Link Prediction on Bipartite Graphs



Link Prediction on Bipartite Graphs

Consider an odd function for the graph kernel

$$A^3$$

Or something more sophisticated

$$\sinh(\alpha A) = \alpha A + \frac{(\alpha A)^3}{3!} + \frac{(\alpha A)^5}{5!} + \dots$$

$$\sinh(\alpha A) = \sum_{i=0}^{\infty} \frac{\alpha^{1+2i}}{(1+2i)!} A^{1+2i}$$

Key Tip

Your choice of link prediction algorithm makes an implicit assumption of the graph kernel – the mechanism for how the graph grows. Different graph types (social, academic, user-item) grow under different mechanisms. Therefore different link prediction algorithms will work better on other graphs.

Objective Function

$$F(\mathbf{A}_{source}) \simeq \mathbf{A}_{target}$$

$$\min ||F(\mathbf{A}_{source}) - \mathbf{A}_{target}||_F$$

Taking the Frobenius Norm

Matrix Decomposition

(Spectral) Matrix Decomposition/Eigendecomposition

$$A = U \Lambda U^T$$

Adjacency Matrices are real, square, and symmetric. We can break them down into Eigenvectors and eigenvalues. The set of eigenvalues is the spectrum of the graph, hence the namesake.

$$F(A) = U F(\Lambda) U^T$$

Easy computation of functions on the Adj. matrix

Decomposition

$$\min ||UF(\Lambda)U^T - \mathbf{A}_{target}||_F$$

$$\min ||F(\Lambda) - U^T \mathbf{A}_{target} U||_F$$

Link Prediction to Regression

$$\min_f \sum_i (f(\Lambda_{ii}) - U_{\cdot i}^T \mathbf{A}_{target} U_{\cdot i})^2$$

Taking the F-norm is analogous to the RMSE between the two matrices. Function F acting upon Lambda only effects the diagonal elements. Therefore trying to minimize the F-norm one can only effect the diagonals of the source matrix to be equivalent (or close) to the right term. Hence the problem is reduced to a curve fitting problem, stated above.

Choosing your Graph Kernel

Table 1 Link prediction functions and their corresponding real functions

	Method	Matrix function	Real function
Section 5.1	Triangle closing	\mathbf{A}^2	$f(\lambda) = \lambda^2$
Section 5.2	Path counting	$p(\mathbf{A}) = \sum_{k=0}^d \alpha_k \mathbf{A}^k$	$f(\lambda) = \sum_{k=0}^d \alpha_k \lambda^k$
Section 5.3	Exponential kernel	$K_{\text{EXP}}(\mathbf{A}) = \exp(\alpha \mathbf{A})$	$f(\lambda) = e^{\alpha \lambda}$
	Neumann kernel	$K_{\text{NEU}}(\mathbf{A}) = (\mathbf{I} - \alpha \mathbf{A})^{-1}$	$f(\lambda) = 1/(1 - \alpha \lambda)$
Section 5.4	Rank reduction	$R_k(\mathbf{A}) = \hat{\mathbf{U}} \hat{\Lambda} \hat{\mathbf{U}}^T$	$f(\lambda) = \begin{cases} \lambda & \text{if } \lambda \geq \lambda_k \\ 0 & \text{otherwise} \end{cases}$

The outer multiplicative coefficient β (see explanation in the text) is not shown, for clarity

Kunegis et al. 2013

Predictions

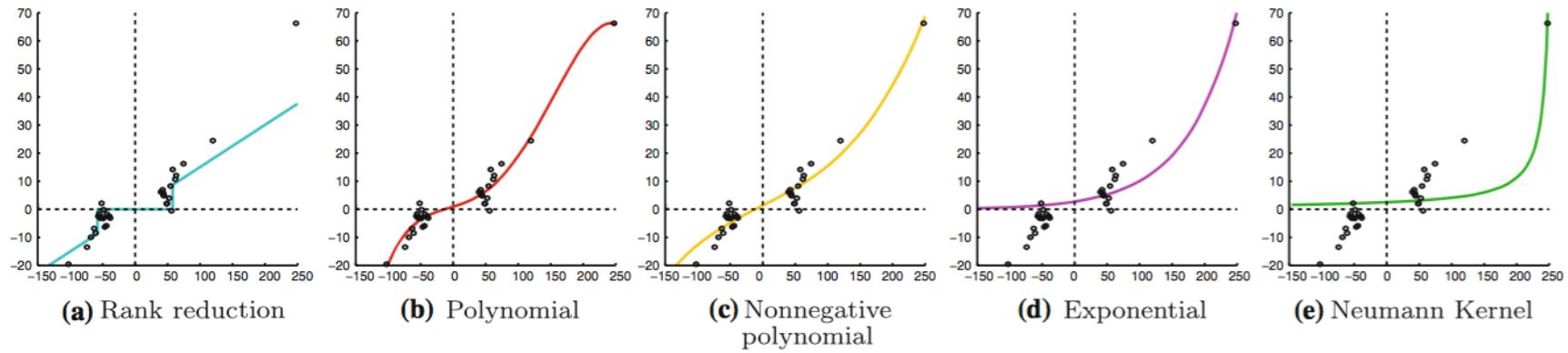


Fig. 11 Curve-fitting individual graph kernels, using the Internet topology network (TO)

Mean Averaged Precision

- For various kernels it is important to check there generalized accuracy, to see which is an appropriate descriptor of the link generation process
- To measure the notion “accuracy”, we use a measurement for ordered lists: Mean Average Precision
- This should be your ‘go to’ when assessing the accuracy of ranked lists.

Average Precision

Average Precision is method to evaluate the precision and *ranking* of a predicted list of retrieved objects

$$AP(q) = \frac{\sum_{k=1}^n [P(k) \times rel(k)]}{\min(m, n)}$$

$$MAP = \frac{\sum_q^Q AP(q)}{Q}$$

Average Precision Example



Relevant links



Ranked list

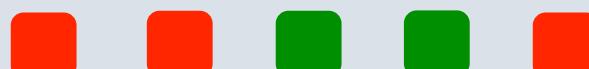
$$AP = \frac{1}{2} (1.0 + 0.67) = 0.84$$

1.0 0.5 0.67 0.5 0.4

Precision



Relevant links



Ranked list

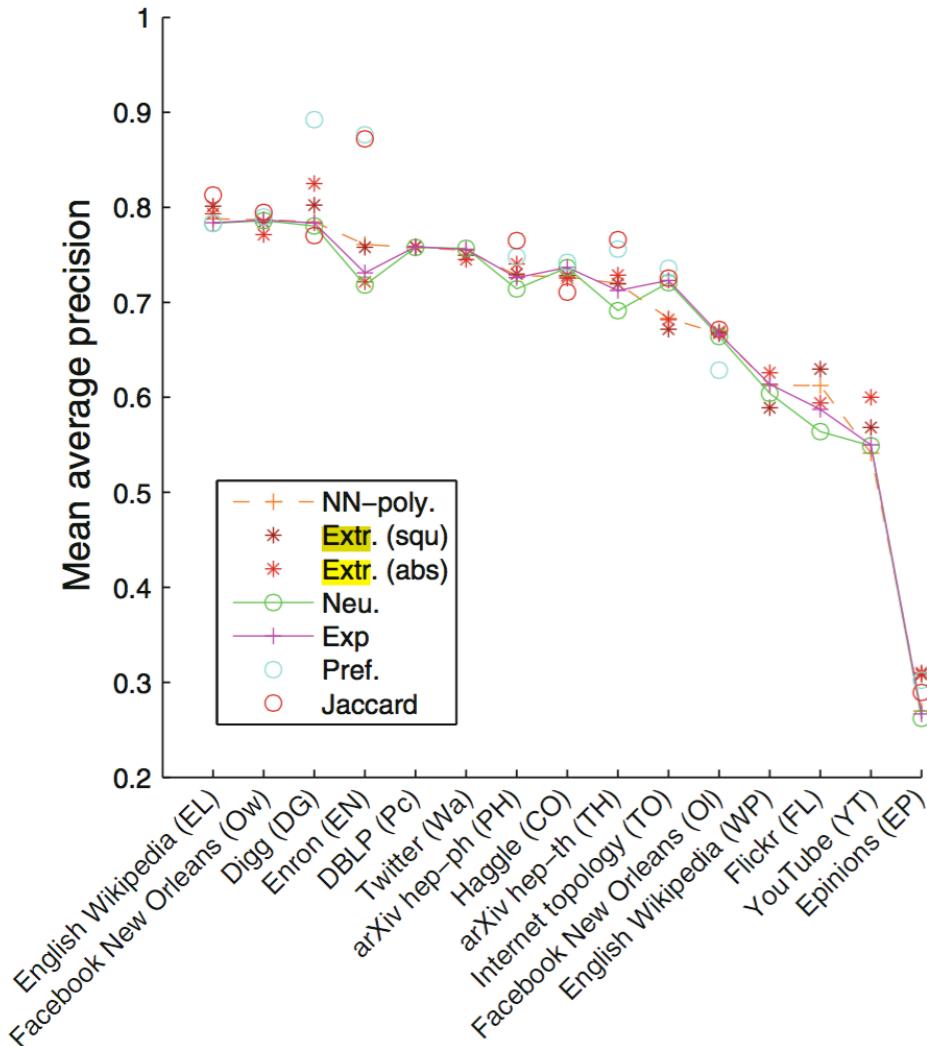
$$AP = \frac{1}{3} (0.33 + 0.5) = 0.28$$

0 0 0.33 0.5 0.4

Precision

$$MAP = (0.28 + 0.84)/2 = 0.68$$

Comparison of Algorithms



Other Methods for Link Prediction

- Random Walks on graphs – Average Commute Time
- Maximum Likelihood methods – Constructing graphs
- Spectral Extrapolation methods – Continuing Eigenvalue trends
- And more... [Lu & Zhou, 2011]

Challenges for Link Prediction

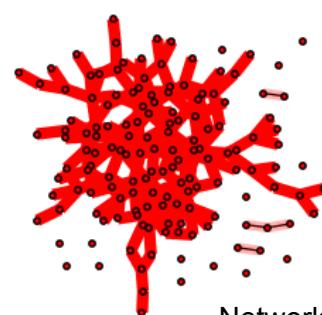
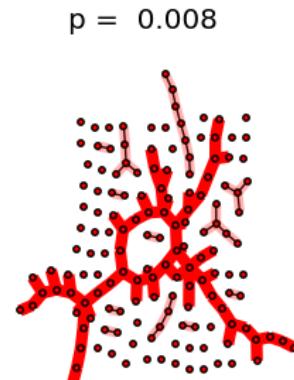
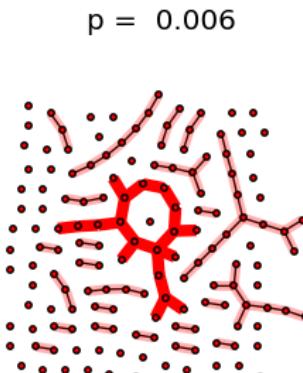
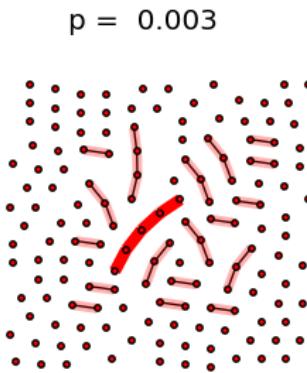
- Cold Start Problem
- Temporal Networks
 - Links can be created and destroyed over time
 - New nodes come in and out of the system
- Link prediction with links of a different nature – e.g., ‘negative’ links.

References

- Link prediction in complex networks: A survey; Lü and Zhou, Physica A; 2011
- The Link Prediction Problem for Social Networks; Liben-Nowell & Kleinberg
- Spectral evolution in dynamic networks; Kunegis et al.; Knowl. Inf. Syst.; 2013
- Online Dating Recommender Systems: The split-complex number approach; Kunegis et al.; ACM 2012
- Mean Average Precision
 - https://en.wikipedia.org/wiki/Precision_and_recall
 - https://en.wikipedia.org/wiki/Information_retrieval
 - Victor Lavrenko: youtube.com/watch?v=pM6DJ0ZZee0
 - https://web-beta.archive.org/web/20110504130953/http://sas.uwaterloo.ca/stats_navigation/techreports/04WorkingPapers/2004-09.pdf

How To: The Setup – quick note

Our algorithms consider only the edges that connect the same set of nodes in the training and test set – usually taken as the nodes of the giant component of the graph



This assumes the graph is static – no new nodes enter the system

By definition: The algorithm will not predict edges for nodes not within the giant component

More about this later

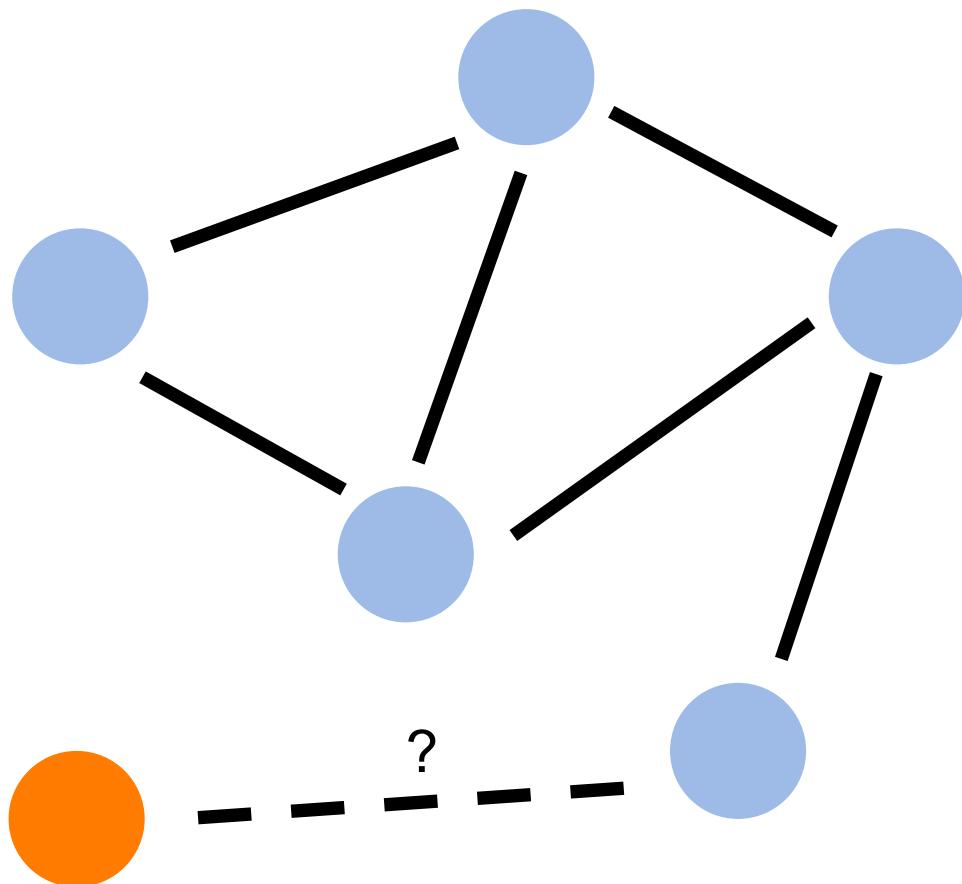
MAP/AP FAQs

- Why do you take the average of the precisions at the relevant documents, and not at all places.
 - My intuition is about the ranking. Relevant docs/links first. So if you take the first example we have and sum over all docs/links retrieved, you get a lower MAP. But suppose you only have 2 relevant docs in a list of five. You rank them first: therefore your AP is 1. Which is what you would want. But if you divide by 5 instead of 2, you get 0.2 accuracy. Which is not in line with what we would want in terms of an accuracy metric for a ranked list.
 - But, giving the precision at a given rank has the effect of weighting the order of the correct guesses. Think: 1 relative doc @ pos 1: AP = 1. @ pos 3: AP = 0.33.

MAP/AP FAQs

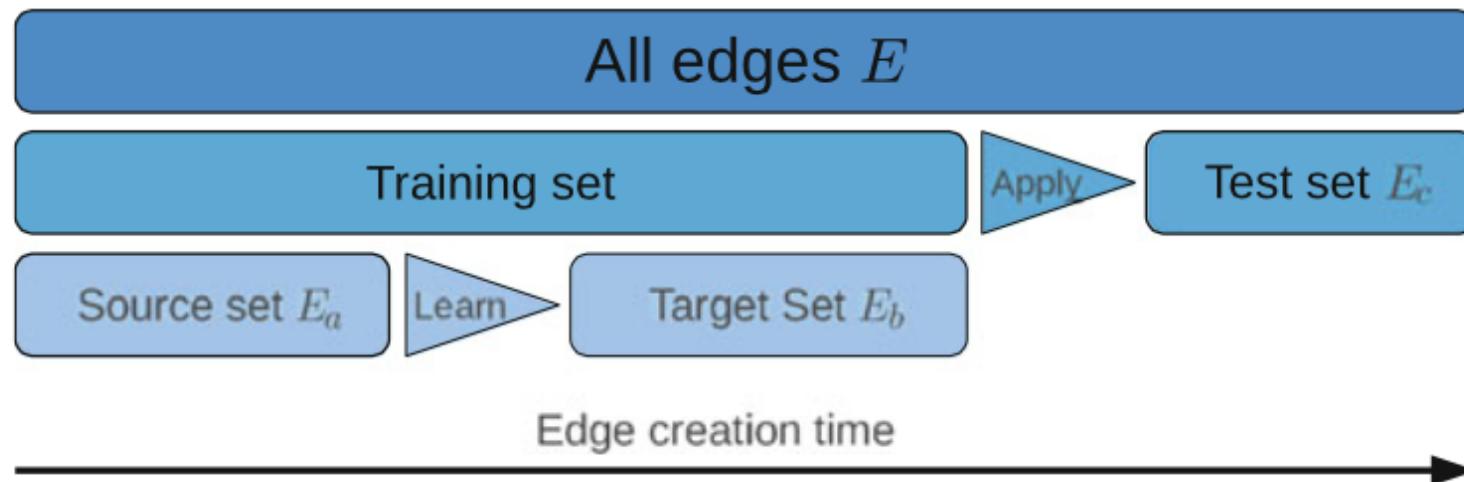
- What if the denominator is zero (e.g., no relevant docs/links)?
 - One simply sets the AP to zero. One could think of this as applying a penalty to the algorithm for retrieving/predicting docs/links when none should be.

The Cold Start Problem – new nodes



Algebraic Methods – How to

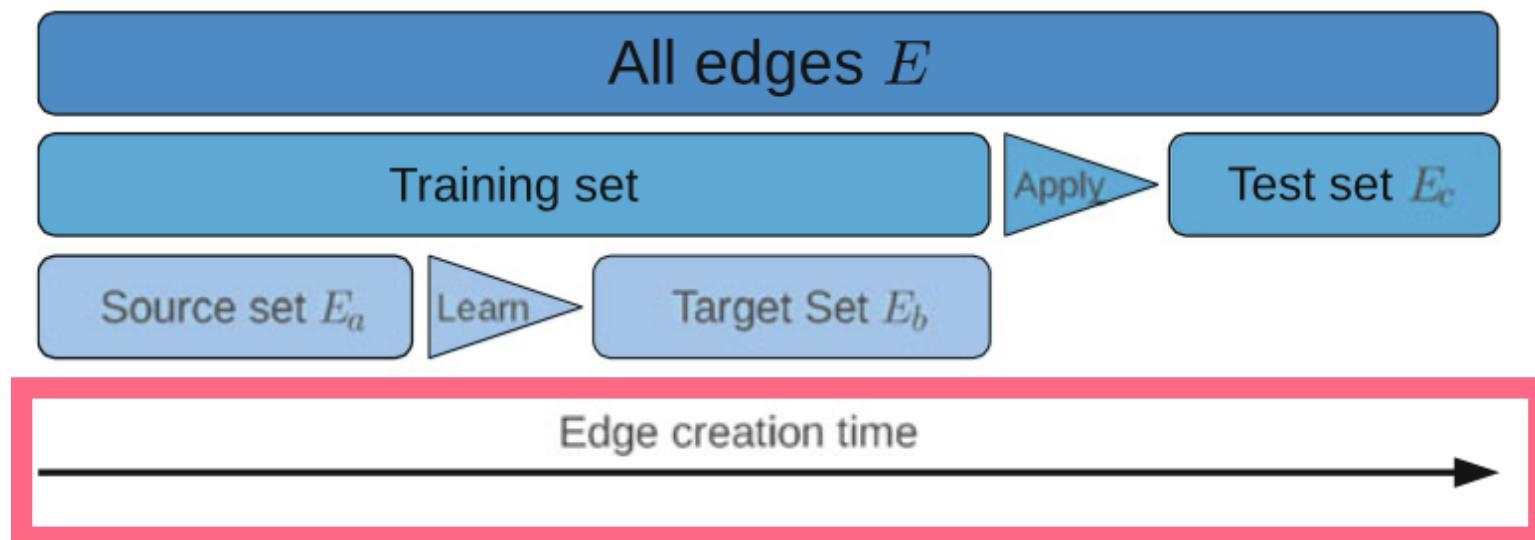
Split data appropriately, giving another split, to allow for learning certain parameters of graph kernels



Kunegis et al. 2013

Algebraic Methods – How to

Split data appropriately, giving another split, to allow for learning certain parameters of graph kernels



Kunegis et al. 2013