University of Glasgow

AI4BioMed

Recommending people in social networks

Algorithmic models and network diversity

Dr. Javier Sanz-Cruzado

WORLD CHANGING GLASGOW

A WORLD TOP 100 UNIVERSITY


ABOUT ME

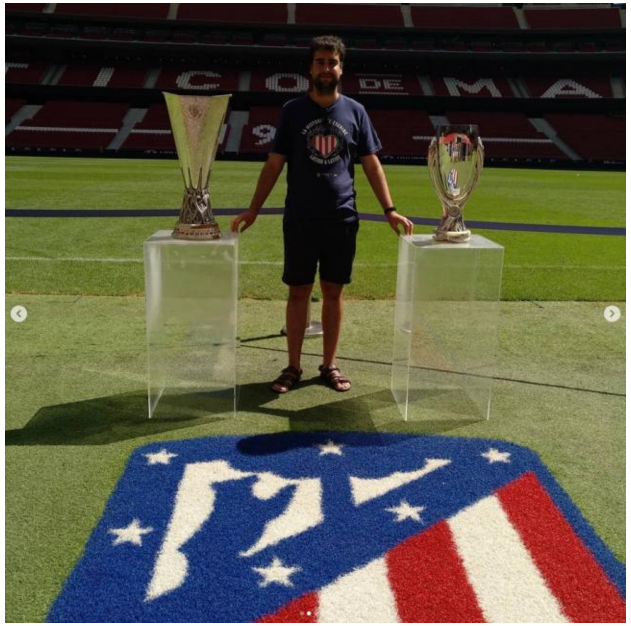
About me

Javier Sanz-Cruzado is a postdoctoral researcher with a strong background in computer science, particularly in the fields of recommender systems, social network analysis, and financial technology. He completed his PhD at the Universidad Autónoma de Madrid, where he focused on algorithmic models for contact recommendations in social networks, earning an outstanding score. His research interests include multi-armed bandits, machine learning, and their applications in recommendation systems.

He has worked on various projects, such as developing an investment recommendation platform for financial advisors in collaboration with the National Bank of Greece. His academic work includes financial recommendation systems and the use of large language models (LLMs) for financial summarization tasks. He has been involved in various international collaborations, publishing papers on topics like stock recommendations and the use of financial knowledge graphs to improve asset recommendations.

Javier has also received several scholarships and awards during his studies, including a PhD scholarship from the Spanish government and excellence scholarships from the Regional Government of Madrid. He is proficient in various programming languages and has taught courses on information search and algorithm analysis(Javier Sanz-Cruzado)(Javier Sanz-Cruzado)(Javier Sanz-Cruzado).

About me



- From Madrid, Spain
- NO. I DON’T SUPPORT REAL MADRID!!
- Things I like:
  - Books (fantasy, sci-fi)
  - Videogames
  - D&D
  - Good food

About me

- People recommendation in social networks
- Multi-armed bandits

UAM
Universidad Autónoma de Madrid

- Financial asset recommendations
- Profitability forecasting

University of Glasgow

- Biomedical NLP

University of Glasgow

2010	2015	2016	2021	2024
Double BSc on Computer Science and Mathematics	Double MSc in Computer Science	PhD in Computer Science	Post-doc Terrier Team	Post-doc AI4BioMed

What is this talk about?

Mild AI

No LLMs were harmed in this research

Graphs

A lot of graphs!

No Biomed

But I hope it gives you ideas in this space

People

Graphs involving people!

In this talk, I will introduce people recommendation in online social networks and issues arising on it

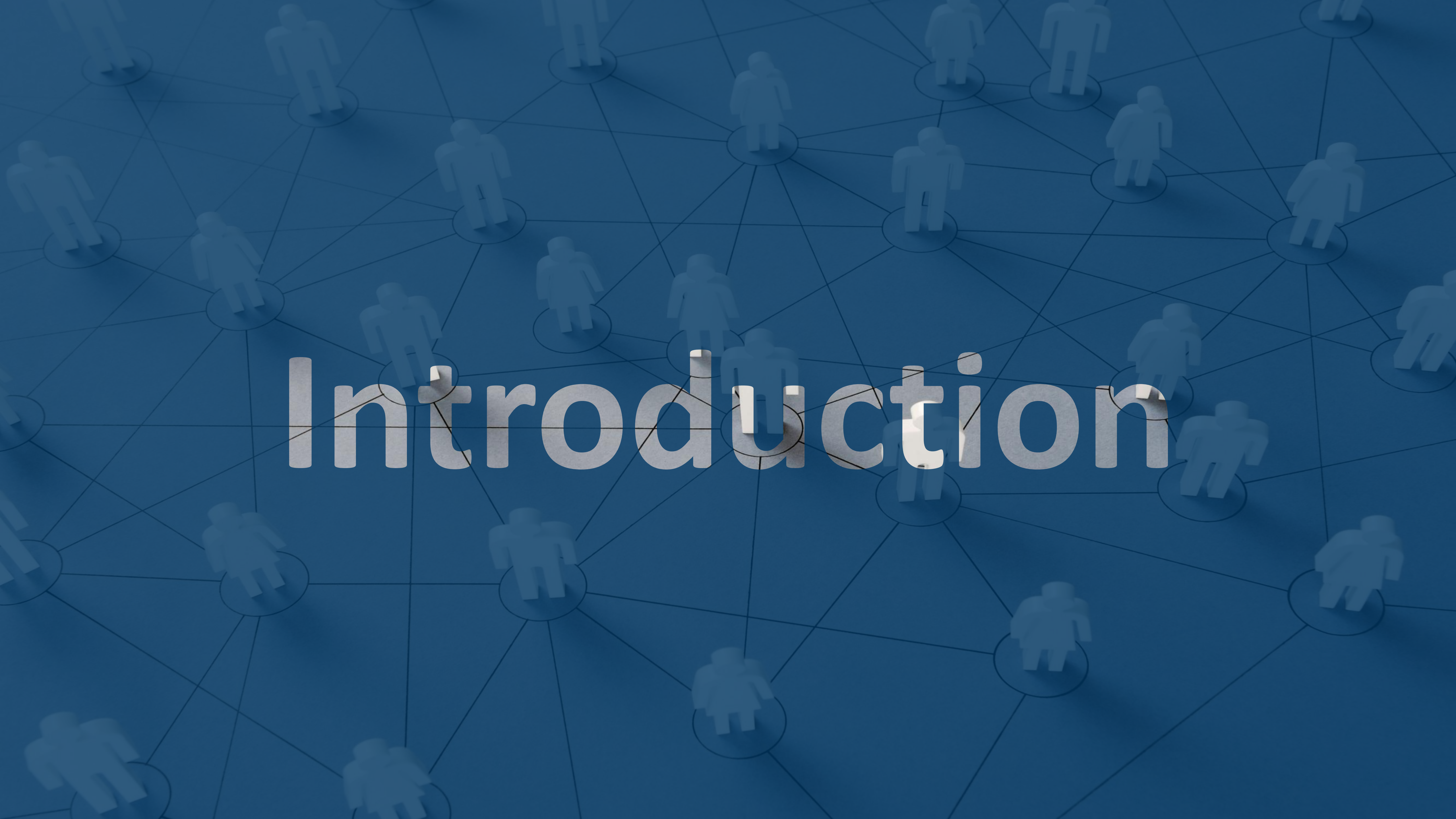
Collaborators

Pablo Castells
Universidad Autónoma de Madrid
Amazon

Sofía M. Pepa
GEOSATIS
Universidad Autónoma de Madrid (formerly)

Iadh Ounis
University of Glasgow

Craig Macdonald
University of Glasgow

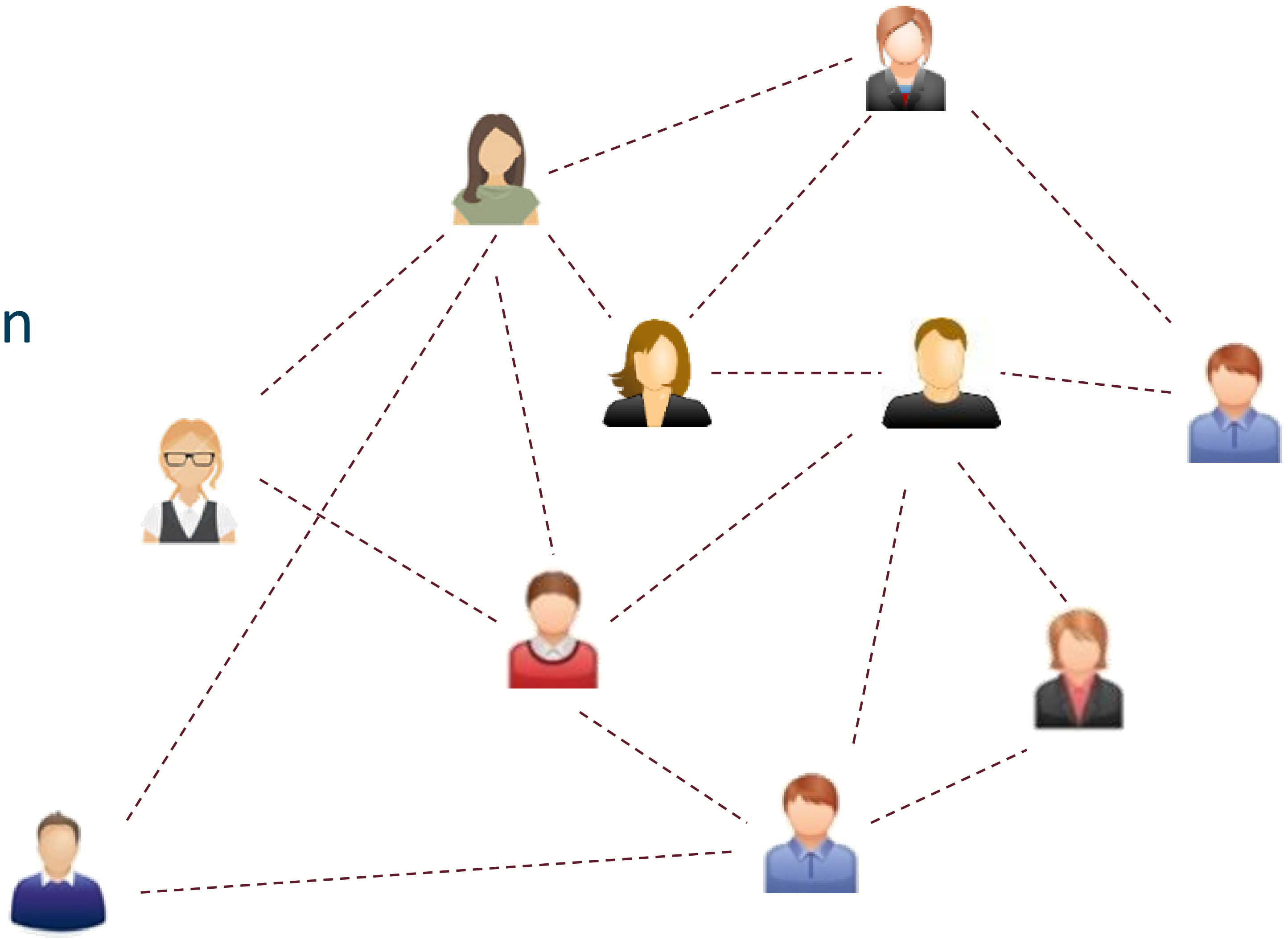

Introduction

University of Glasgow

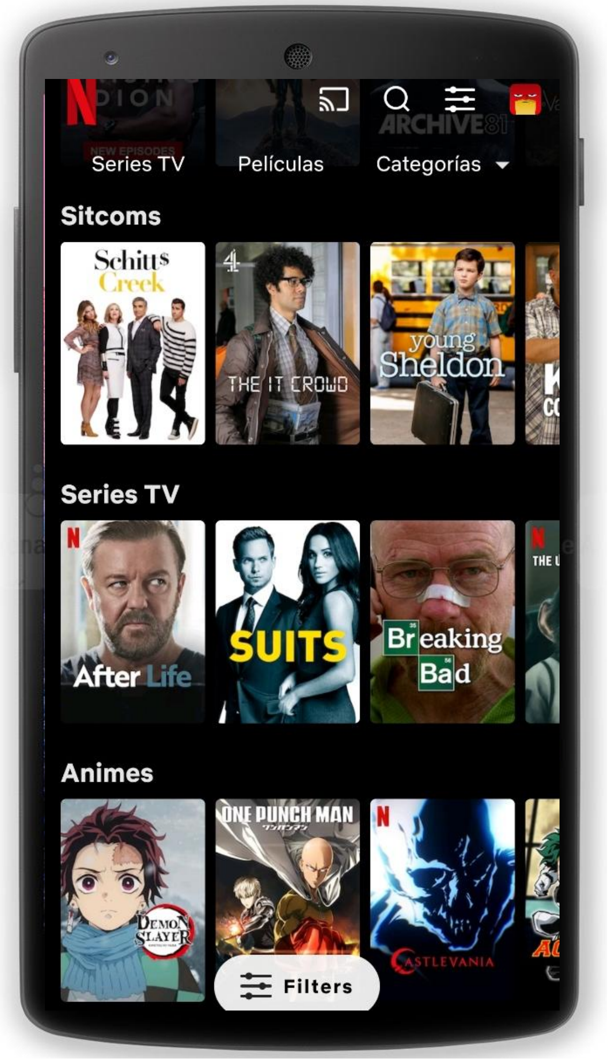
Online social networks

Online social networks

- Establish new connections
- Communication
- Share and receive information
- Changes to our society
  - Politics
  - Privacy
  - Lifestyle
  - Communication



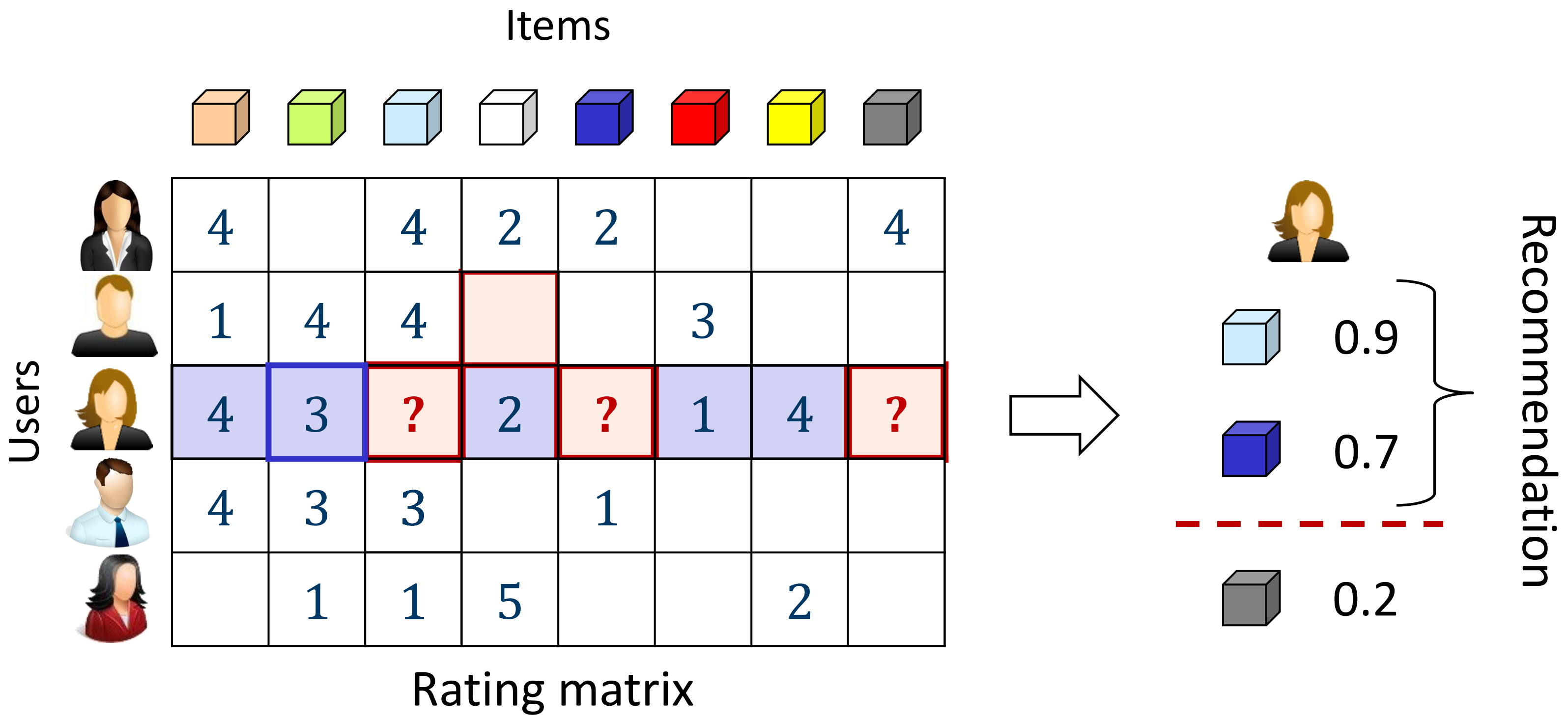
Recommender systems



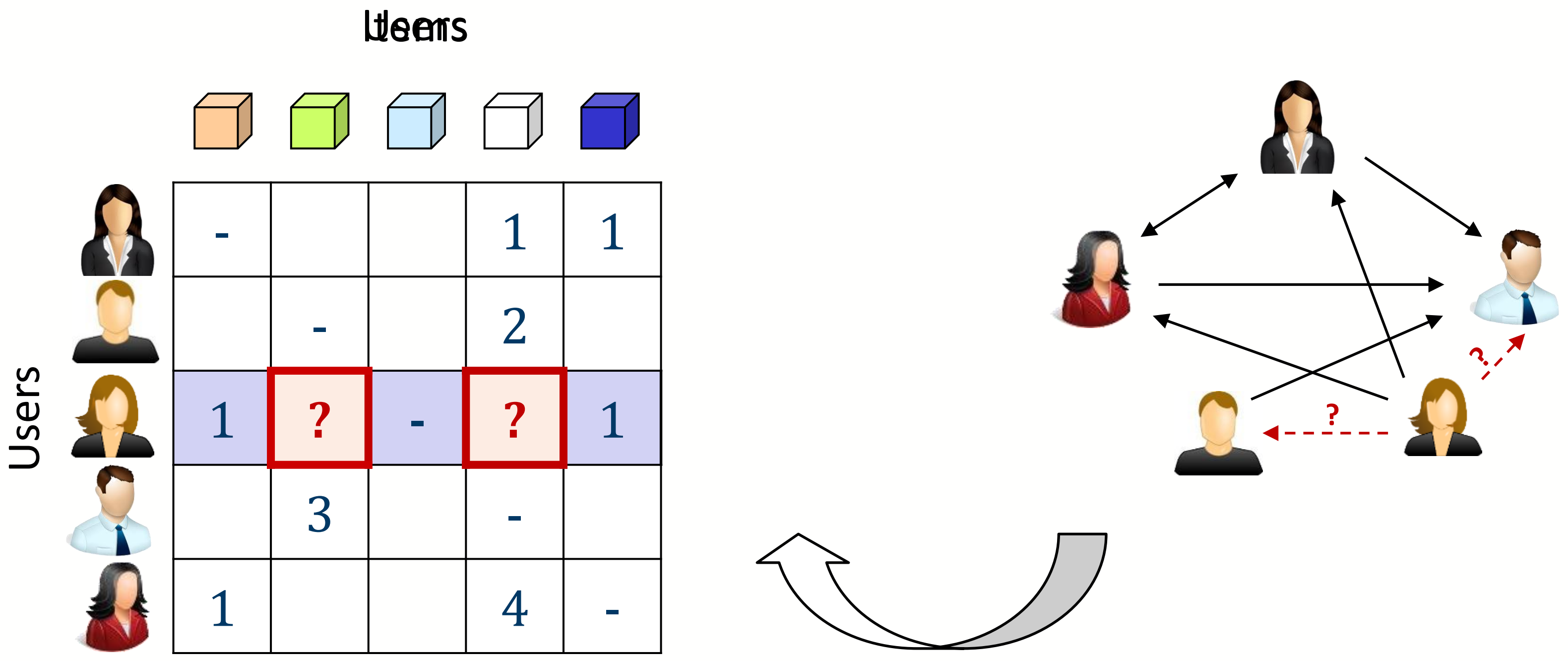
- **Goal:** From past user interactions, suggest items they might be interested in.

- **Multiple domains**
  - Audiovisual content: Netflix, Spotify
  - E-commerce: Amazon, eBay
  - Academic publications: Google Scholar, Mendeley
  - **Social networks: Twitter, Facebook**

The recommendation task



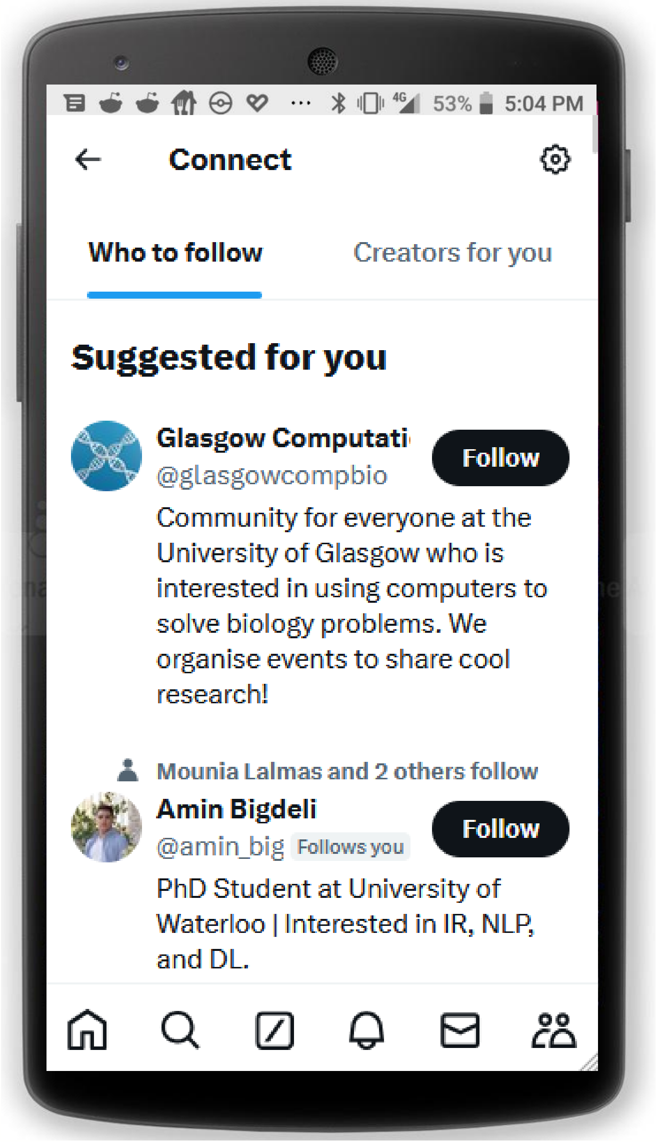
Contact recommendation

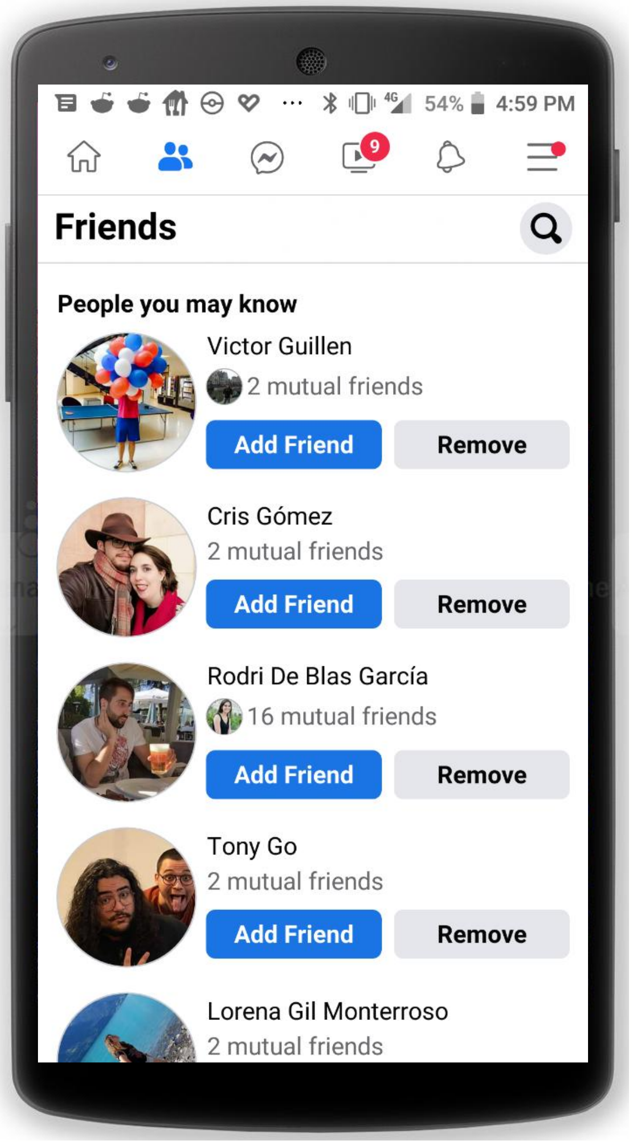


- Items = users
- Availability of social relationships
- Rating matrix = adjacency matrix

University of Glasgow

Contact recommendation examples





Why contact recommendation?

- Particular characteristics
  - Development of new methods
  - Use of social network analysis

- Creation of new links
  - Main asset of online social networks
  - Communication channels
  - Source of information
  - Increase engagement of users

This presentation

Part I: Algorithmic models

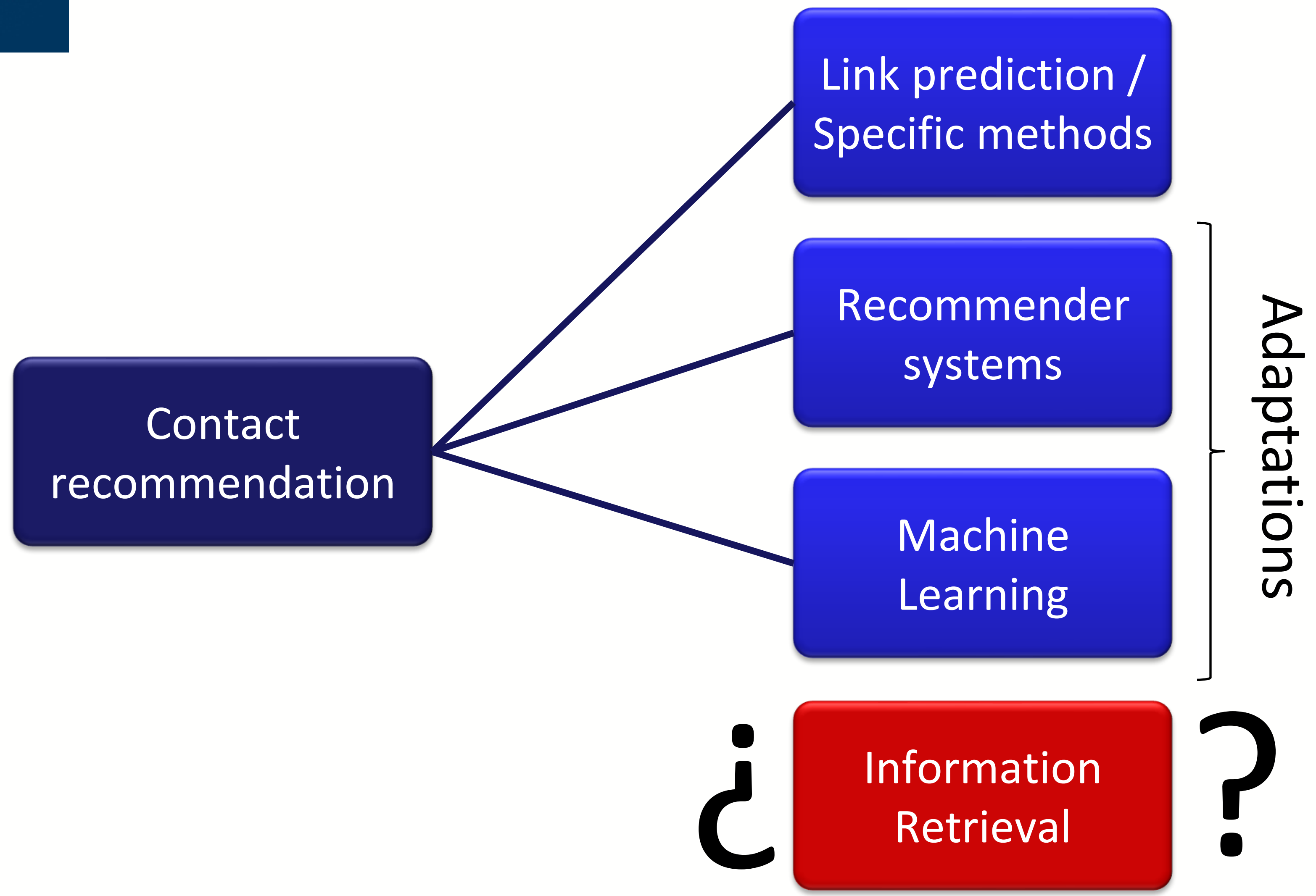
Explore the adaptations of text information retrieval (IR) models to the contact recommendation task.

Part II: Network diversity

Study the effect of contact recommendations on the properties of social networks.

Algorithmic Models

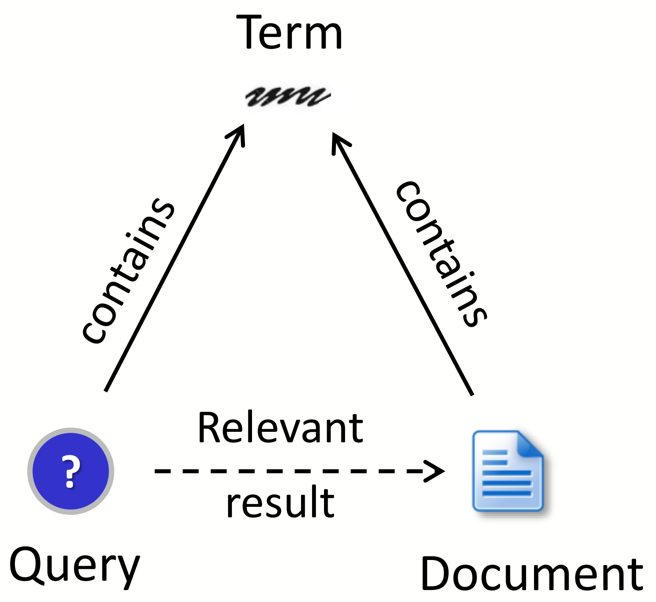
Motivations



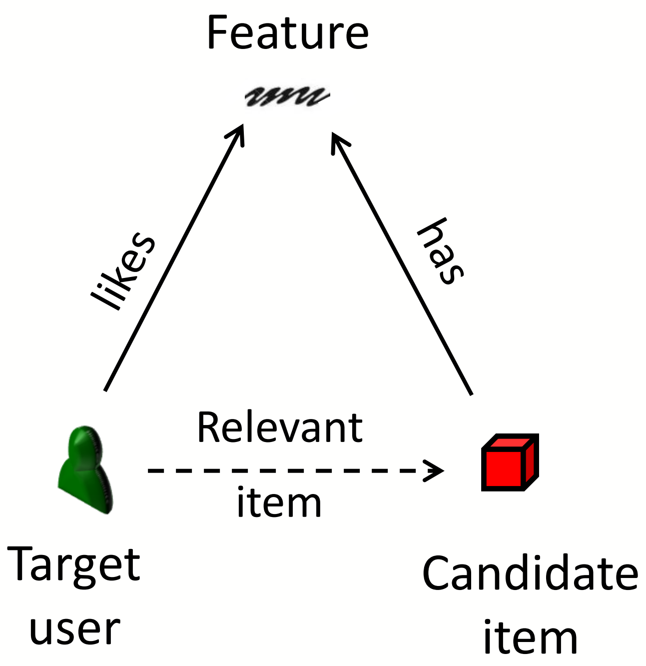
(Sanz-Cruzado & Castells 2018)

IR vs. Recommendation

Text IR

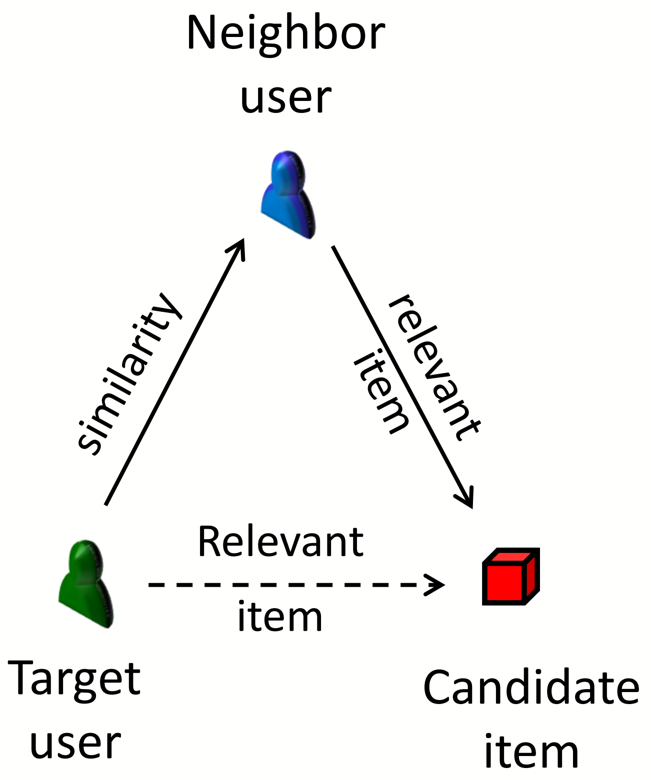


Content-based Recommendation



(Adomavicius & Tuzhilin 2005)

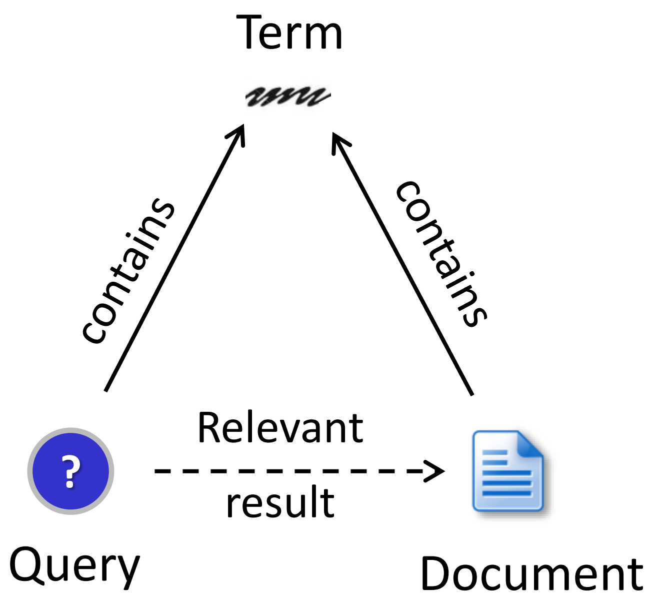
Collaborative filtering



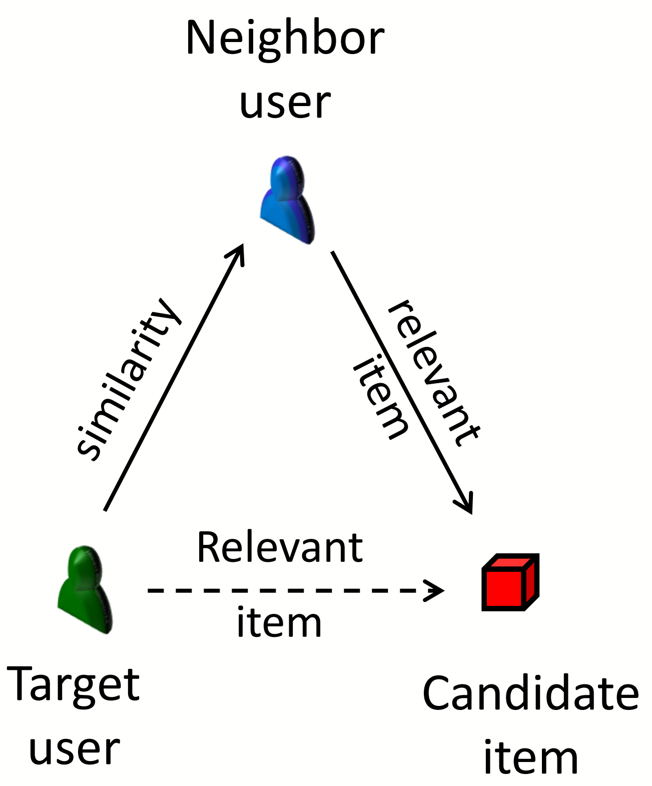
(Bellogín et al., Parapar et al. 2013, Wang et al. 2008, Valcarce et al. 2017)

IR vs. Contact recommendation

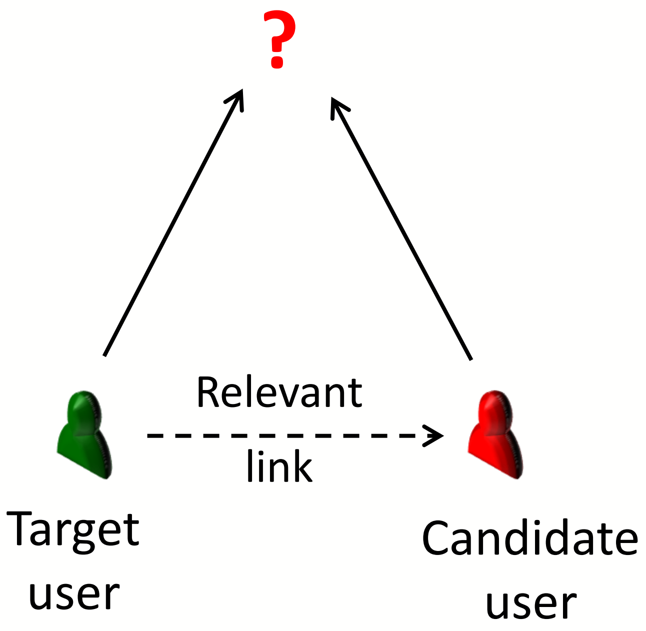
Text IR



Collaborative filtering

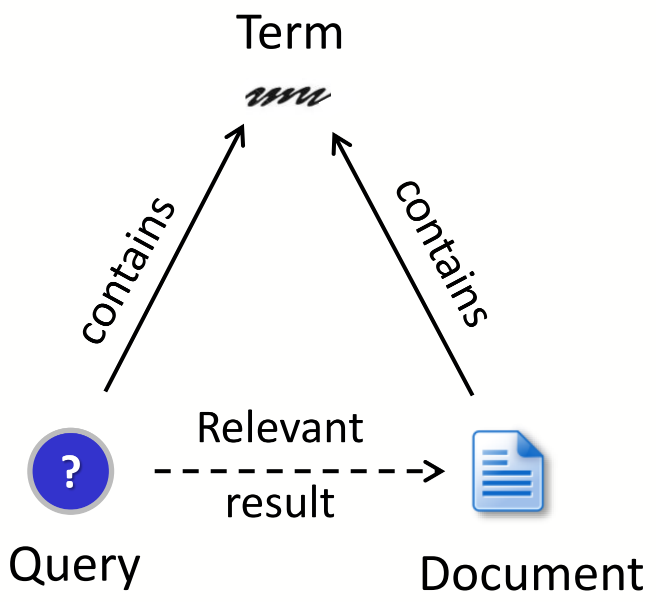


Contact recommendation

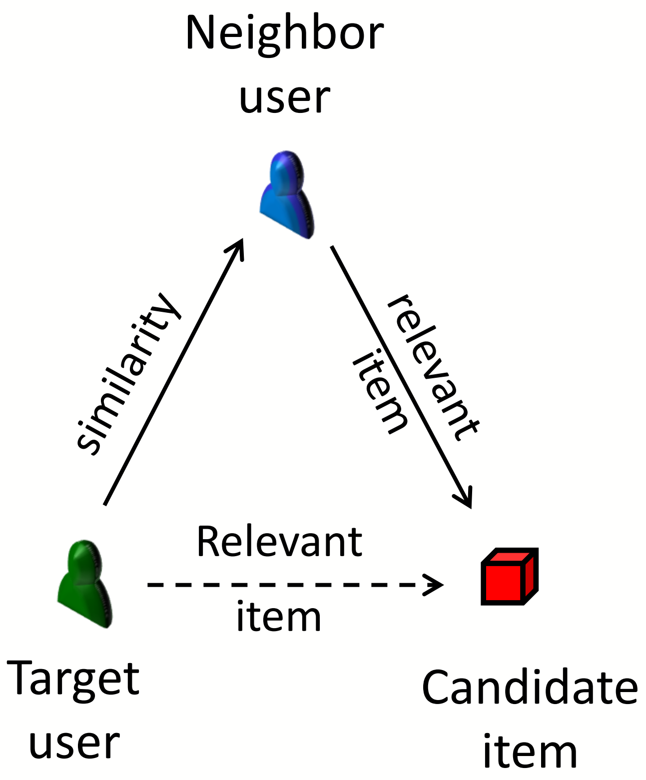


IR vs. Contact recommendation

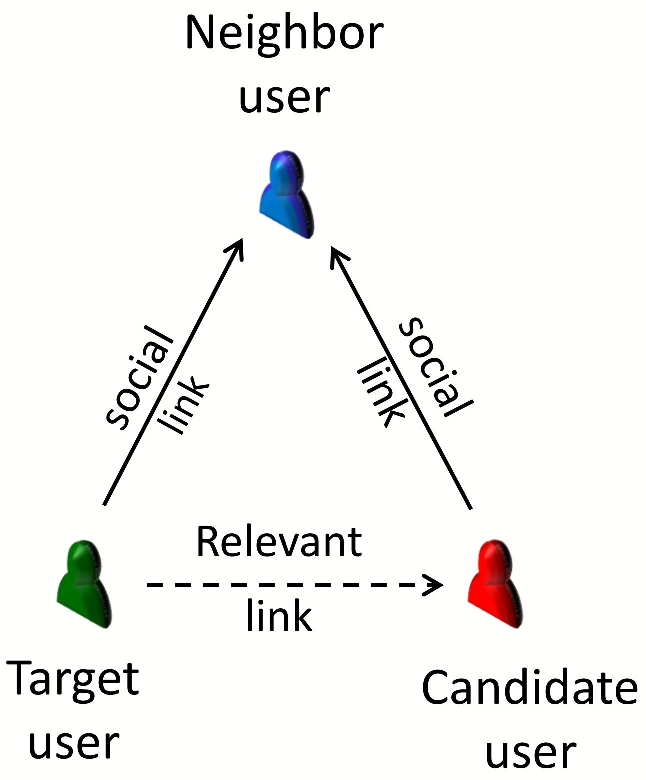
Text IR



Collaborative filtering

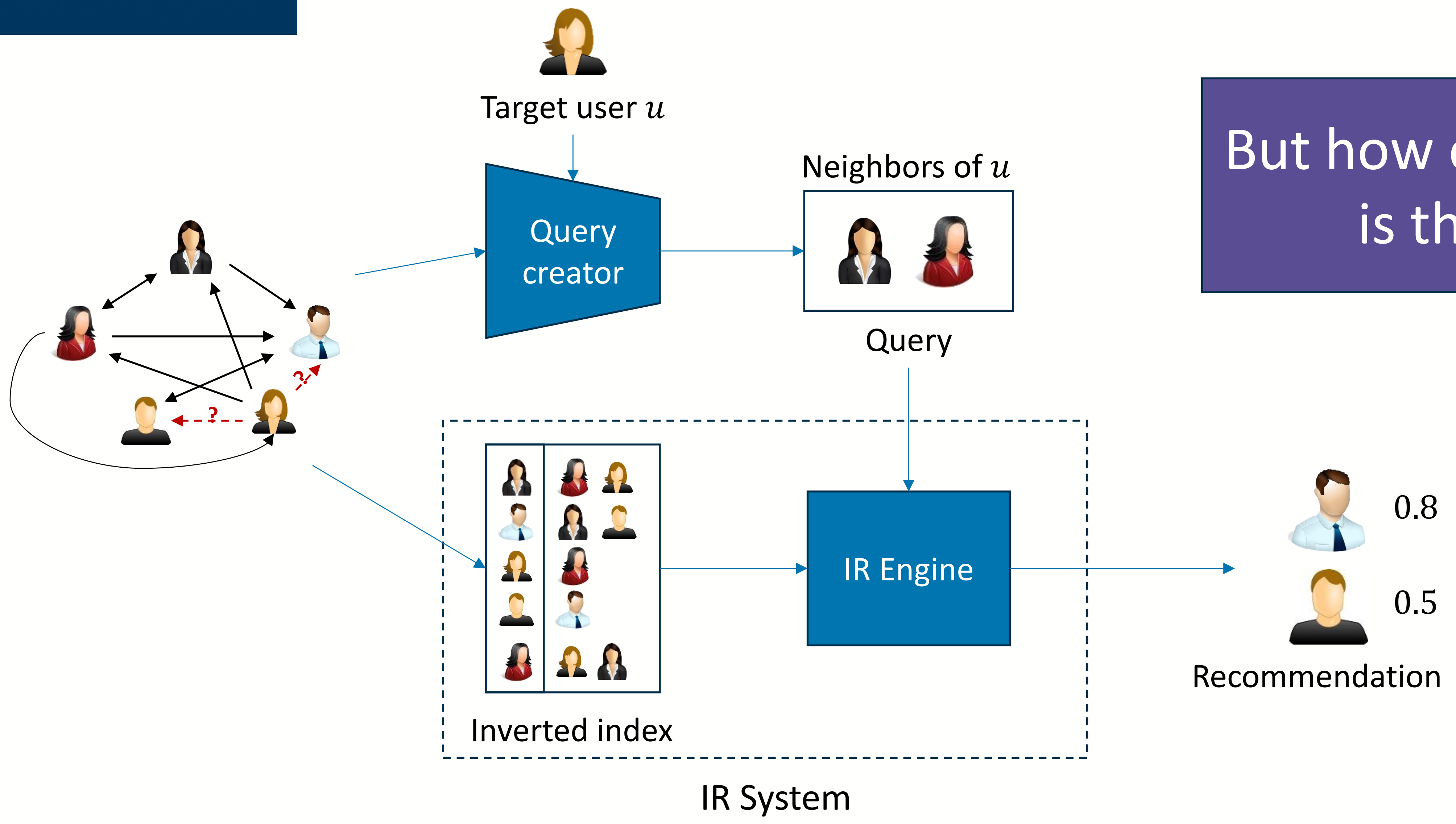


Contact recommendation



(Hannon et al. 2010)
(Sanz-Cruzado et al. 2020)

Use of IR models for contact recommendation



But how effective is this?

More details on Slides 58-59!

University of Glasgow

Data

REST API

Snowball sampling over mentioned / retweeted users

1 month

All tweets between
19th June 2015 – 19th July 2015

200 tweets

Last 200 tweets
of collected users

Interactions

Implicit network

u has a link to v
if u retweets, mentions v

Follows

Explicit network

u has a link to v
if u follows v

Pre-existing dataset

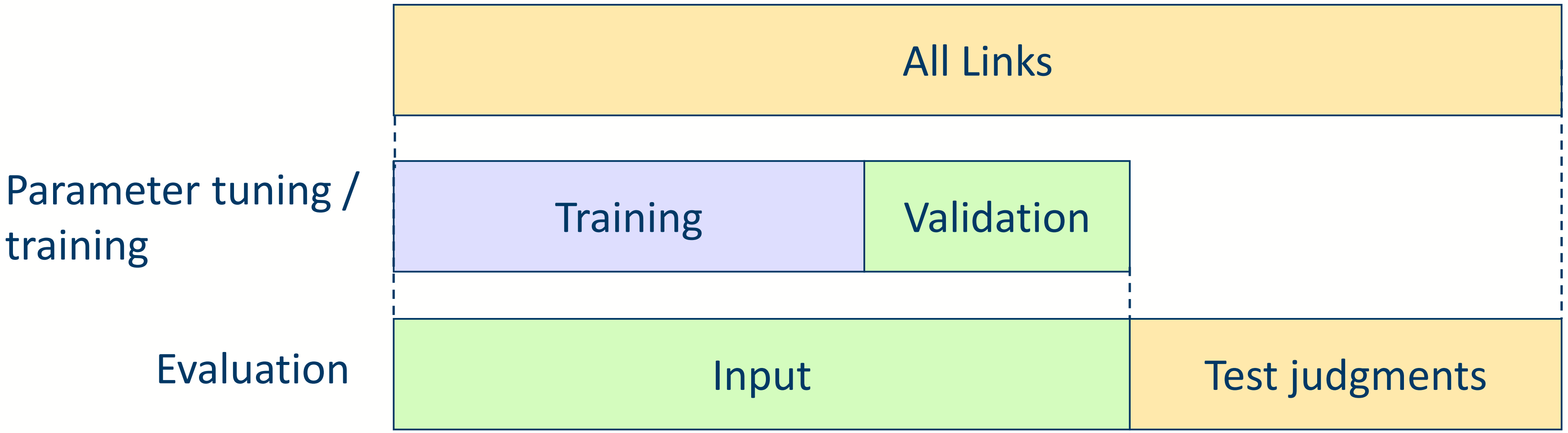
Facebook

Union of 10 ego-networks

SNAP

Methodology

- Split:



- Hyperparameter selection: grid search (nDCG@10)
- Evaluate using IR metrics on test: nDCG@10, MAP@10

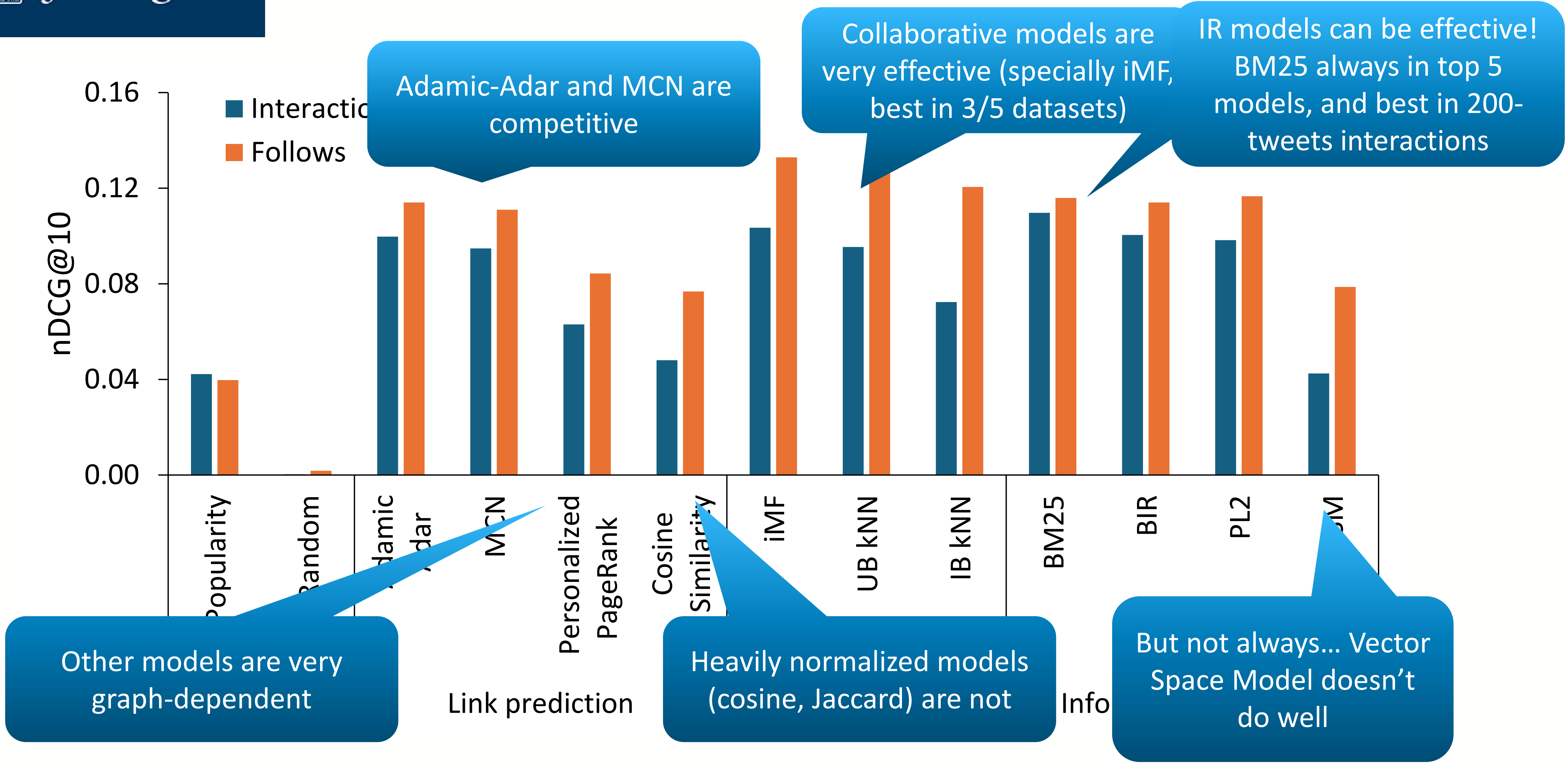
Dataset statistics

	Twitter 1-month		Twitter 200-tweets		Facebook
	Interactions	Follows	Interactions	Follows	
Users	9,528	9,770	9,985	9,964	4,039
Input edges	170,425	645,022	104,866	427,568	56,466
Test edges	54,335	81,110	21,598	98,519	17,643
Directed	✓	✓	✓	✓	✗
Weighted	✓	✗	✓	✗	✗
Split type	Temporal	Temporal	Temporal	Temporal	Random
Density	0.0018	0.0067	0.0013	0.0048	0.0087

Algorithms

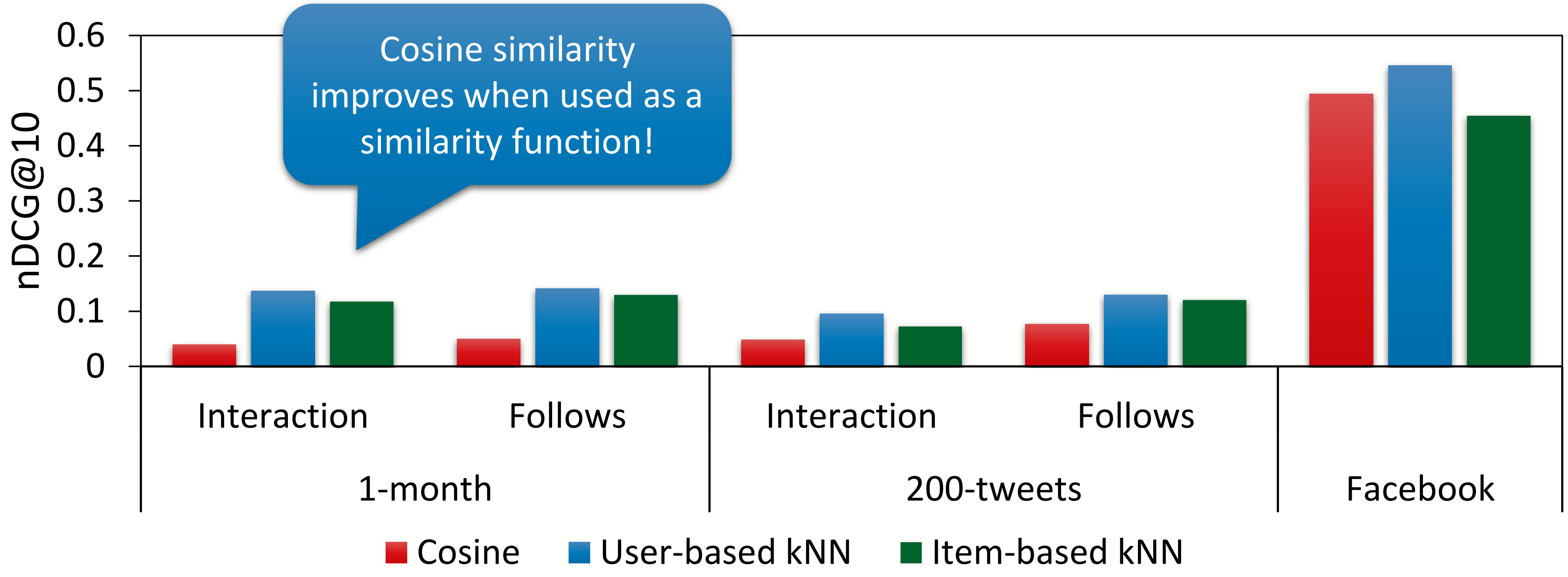
- **IR models:**
  - **Probability ranking principle:** BM25, BIR, ExtremeBM25
  - **Language models:** Query likelihood (QLJM, QLD, QLL)
  - **Divergence from randomness:** PL2, DFRee, DFReeKLIM, DLH, DPH
  - Vector space model (VSM)
- **General collaborative filtering**
  - User-based / Item-based kNN (cosine similarity)
  - Implicit matrix factorization (iMF)
- **Specific approaches**
  - **Friends of friends:** Adamic-Adar, MCN, Jaccard, cosine similarity
  - **Random walks:** Personalized PageRank, Money,…
  - **Path-based:** Local Path Index, Katz…
- **Sanity check:** Random and most popular

Results (nDCG@10) – 200-tweets dataset



Can we do better?

- User-based and Item-based kNN used cosine similarity
- Cosine similarity was also a standalone model

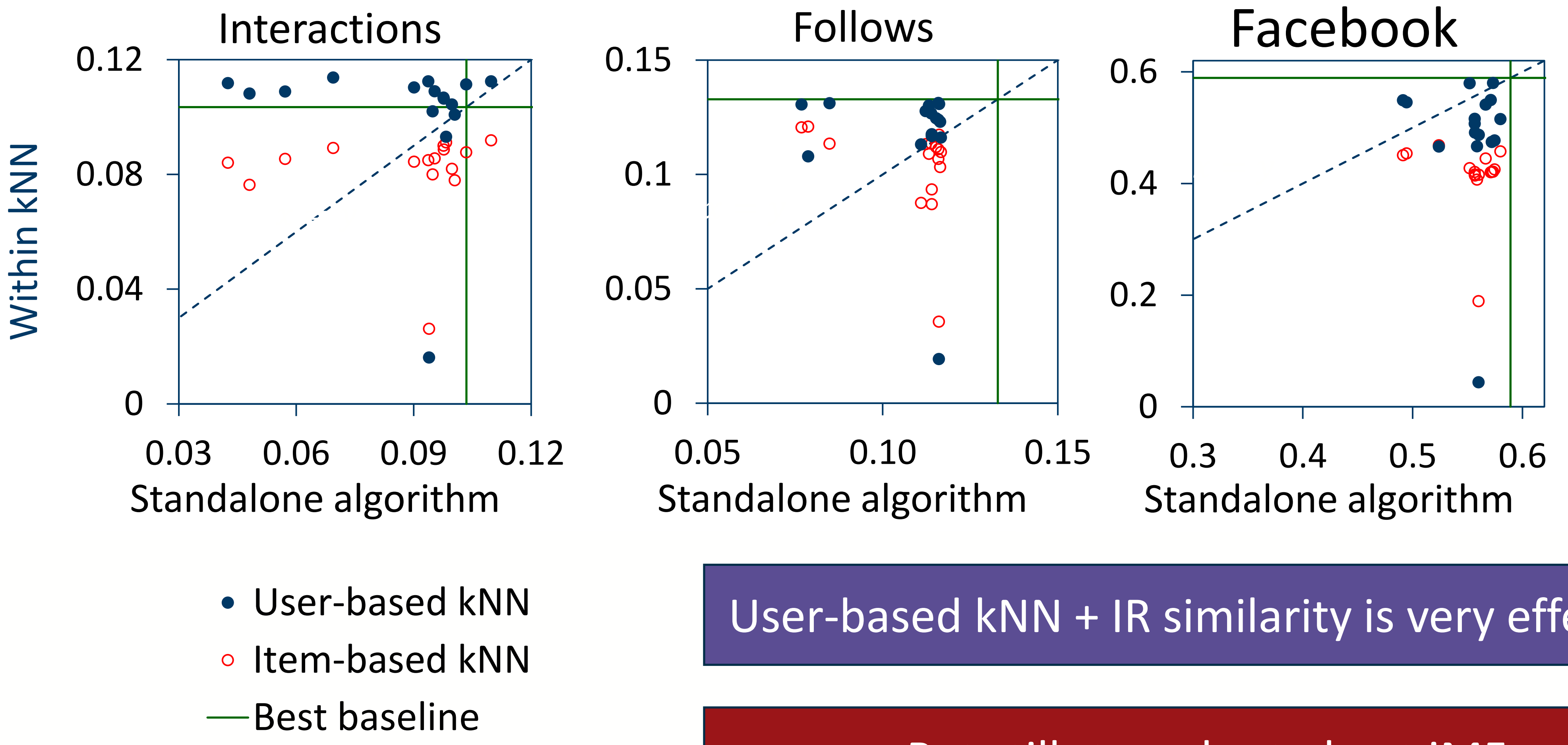


What if we try the same with IR models?

More details on how on slide 60!

Results kNN + IR (nDCG@10)

Twitter 200-tweets



User-based kNN + IR similarity is very effective

But still struggles to beat iMF

University of Glasgow

Can we do even better?

- **Idea:** Learning to rank (Liu 2007)
    - Supervised machine learning models
    - Very effective in IR

- **How does it work?**
    - Sample candidates
    - Generate features for each target-candidate user pair
    - Generate recommendation ranking

Our experiments

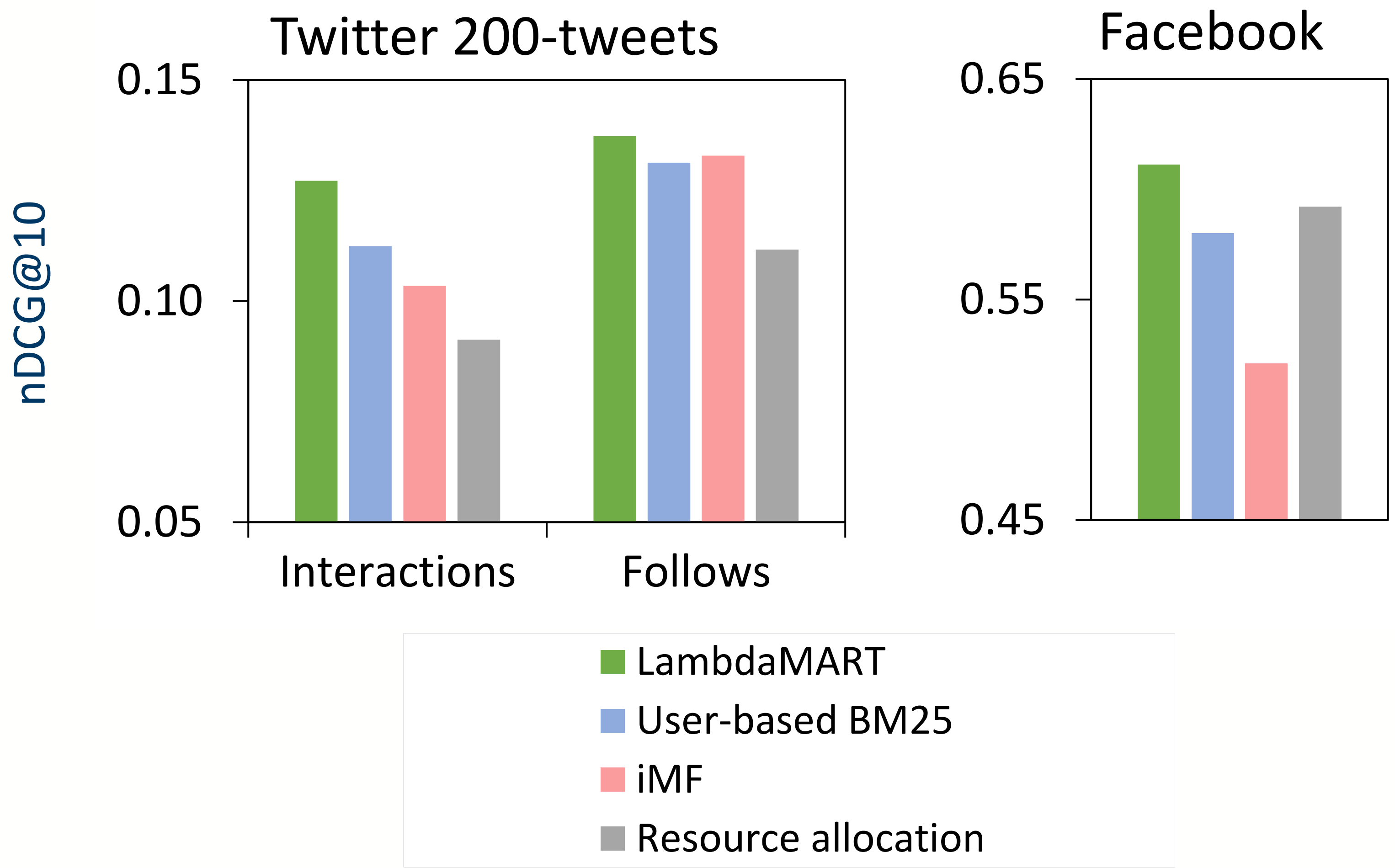
- **LETOR algorithm:** LambdaMART (Burges 2010, Ganjissafar et al. 2011)

- **Features:** Scores of contact recommendation methods
  - IR models
  - Friends of friends (FOAF) approaches
  - User-based / Item-based kNN + IR / FOAF

- **Sample suitable candidates:** use IR models

Results (nDCG@10)

LambdaMART improves best recommendation baselines



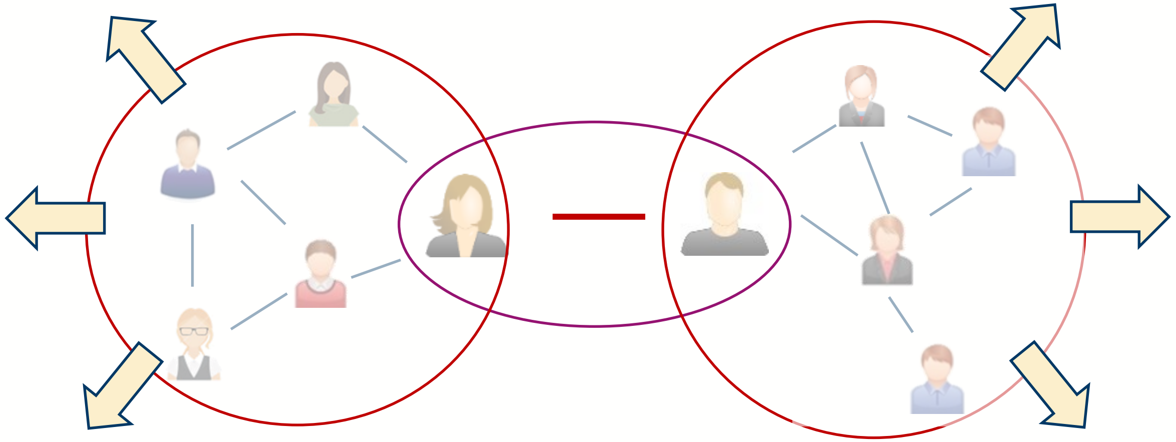
Conclusions

- We can use IR models as contact recommendation algorithms
- Direct IR models are both effective and efficient (BM25)
- IR-based models are better as neighborhood selectors for kNN
- Learning to rank techniques improve the accuracy of best state of the art algorithms
- IR models are effective in three different roles in contact recommendation
  - Direct recommenders
  - Neighborhood selectors in kNN
  - Samplers and features in learning to rank


Network diversity

University of Glasgow

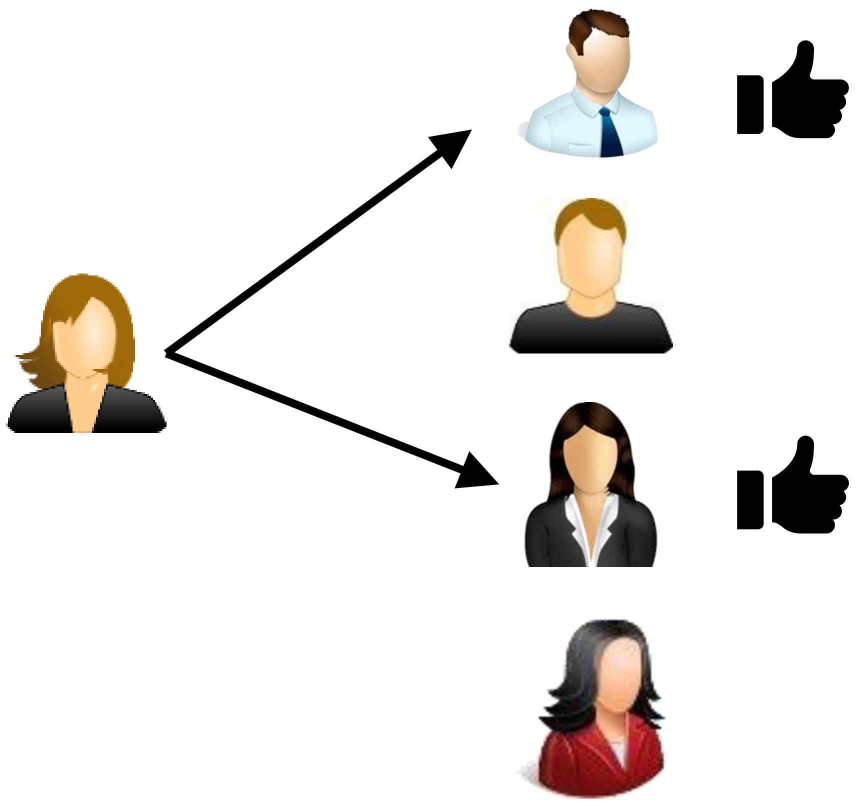
Beyond accuracy



- Users in the network are not isolated
- A few links can cause global effects
- Different links – different effects
- Contact recommendation
  - 500 million new links/month on Twitter (as of 2015)
  - Potential to drive network evolution

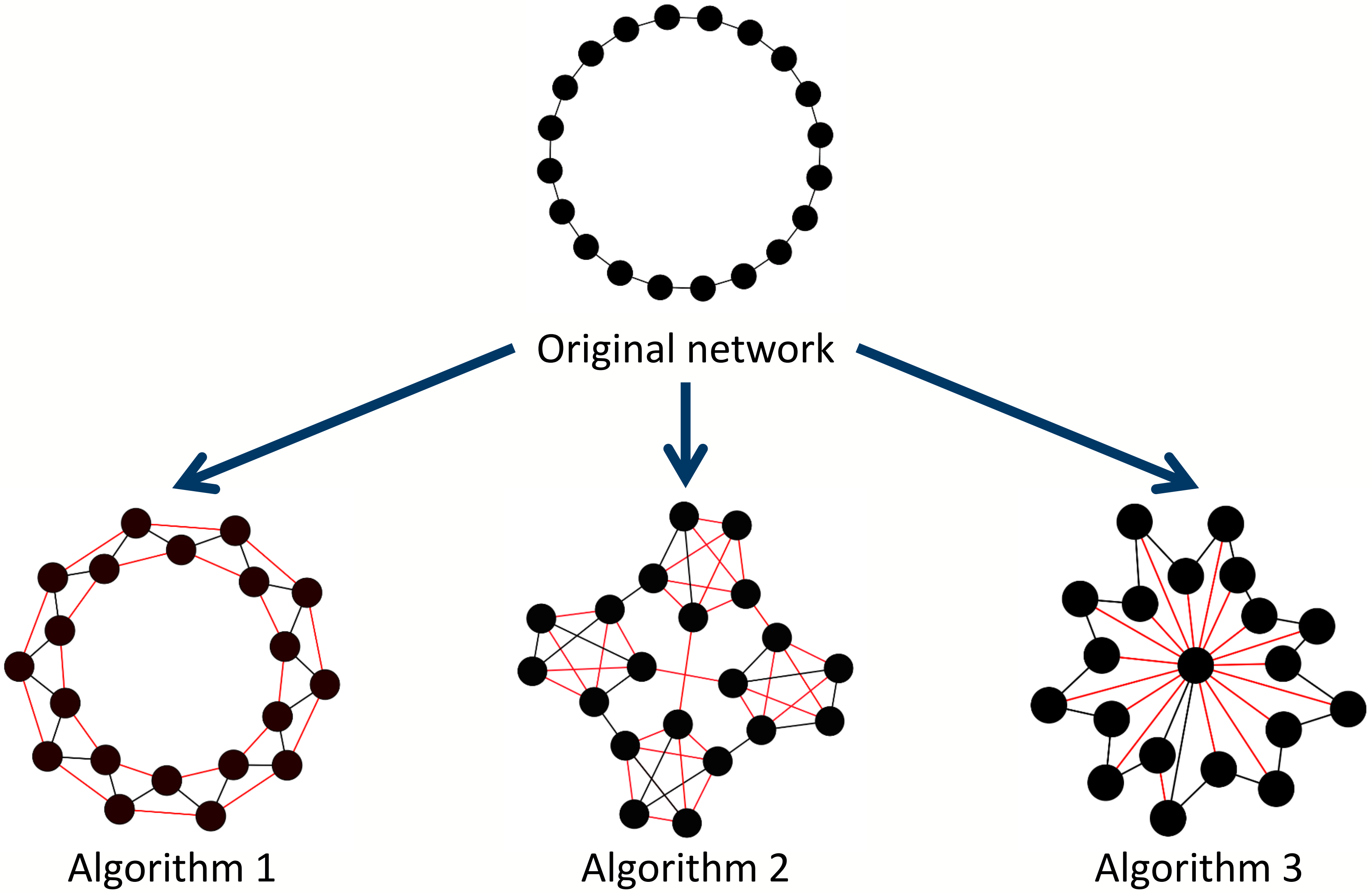
Accuracy

- Fundamental goal of contact recommendation
- Increase network density
- Limitations:
  - **Local perspective:** average over isolated users
  - **Narrow perspective:** one-dimensional utility



University of Glasgow

Effects on network structure



Goals

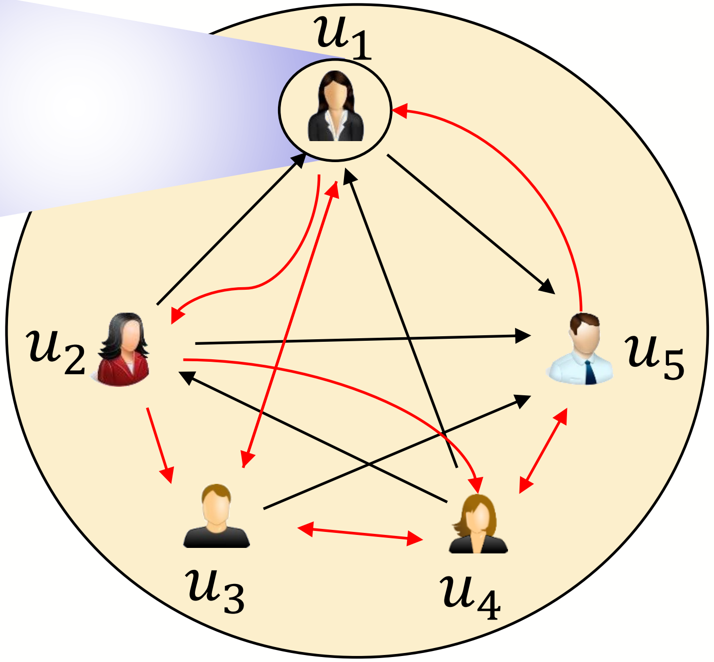
Define suitable metrics for measuring the effects of contact recommendation beyond accuracy.

Determine their meaning for the users in the network.

How to measure?

User	Score
u_2	0.9
u_3	0.8
u_4	0.1

Recommendation ranking



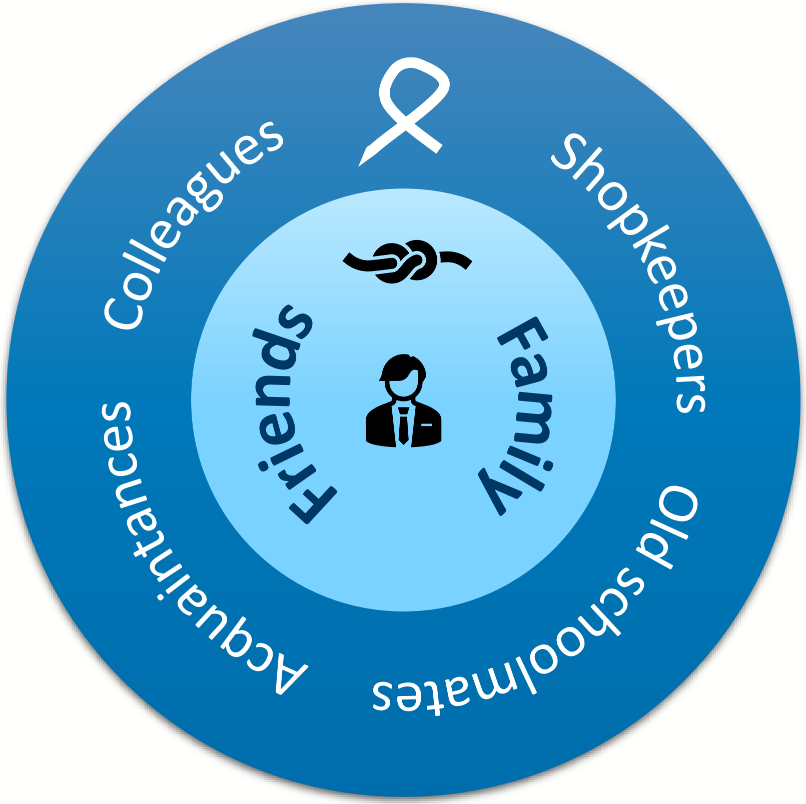
Structural metric

Potentially relevant structural features

- **Structural diversity**
  - Source of novel information.
  - Enrichment of the information flow.
  - Related to the notion of weak tie (Granovetter, 1978)

- **Strength of a tie**
  - Measures the involvement of users in the tie.
  - **Strong ties:** family, close friends.
  - **Weak ties:** people you meet in conferences, shopkeepers.

- In the network structure: non-redundant links



Weak links: local notions

- Consider the direct environment of the link.

- **Triadic closure:** minimum unit of structural redundancy.

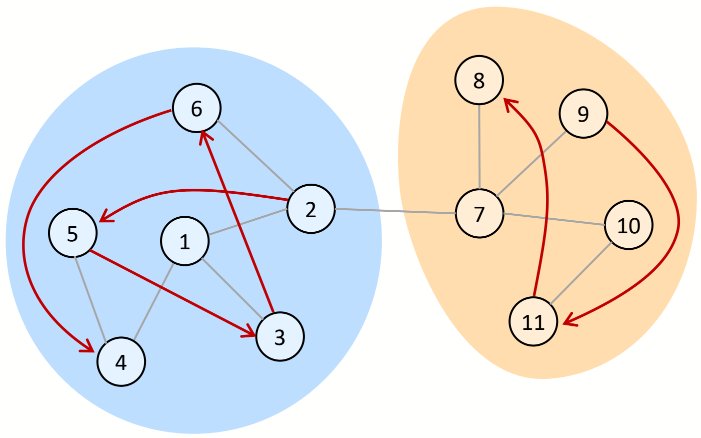
a) Non-redundant triad b) Redundant triad

- **Metric:** clustering coefficient complement.

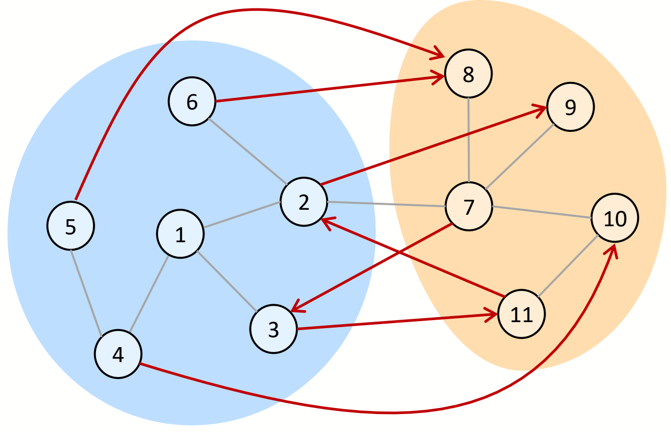
Measures the proportion of non-redundant triads in the network.

Weak ties: global notions

- **Weak ties:** links between communities (De Meo et al. 2012)
  - Tightly connected groups of nodes
  - Few connections outside the group
- **Modularity complement (MC):** number of weak ties

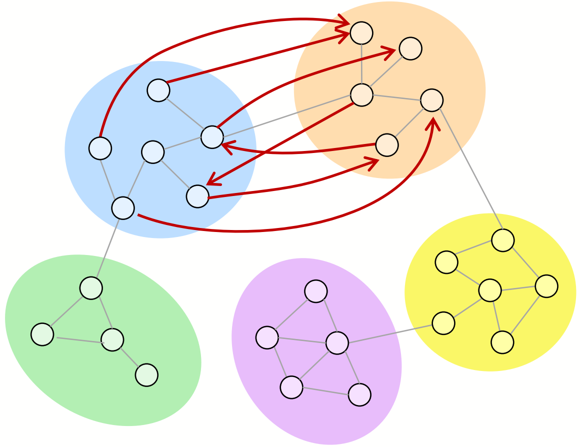


Low MC

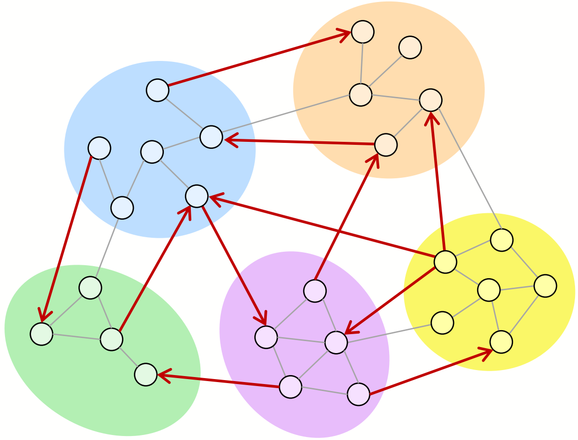


High MC

Weak ties: global notions (II)



Weak-link redundancy



Weak-link diversity

- **Community edge Gini complement (CEGC)**
    - **New metric**
    - Distribution of weak links between pairs of communities
    - Based on the Gini index

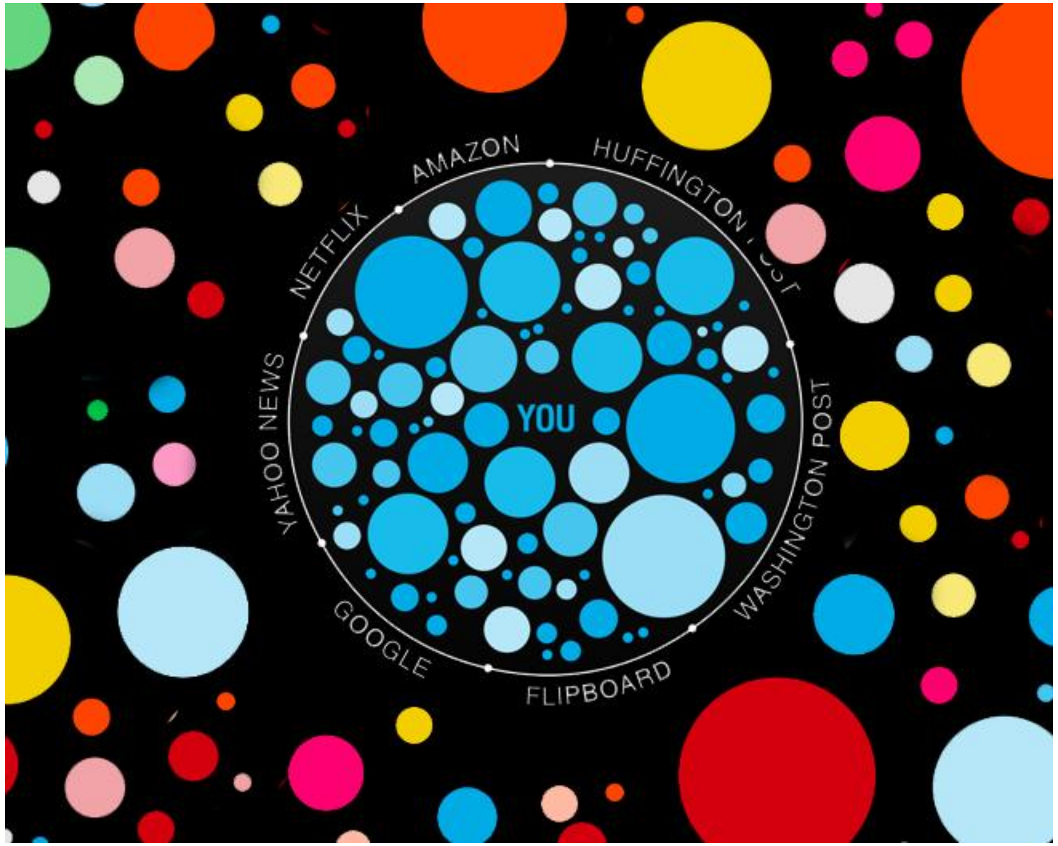
Effect of different recommenders

Algorithm	nDCG@10	Clustering coefficient	Modularity	Community Gini
iMF	**<u>0.139</u>**	0.902	0.155	0.045
BM25	0.104	0.878	0.150	0.041
Adamic-Adar	0.098	0.882	0.149	0.041
MCN	0.092	0.879	0.145	0.040
Pers. PageRank	0.100	0.915	0.182	0.054
Popularity	0.057	0.924	**<u>0.295</u>**	0.061
Random	0.001	**<u>0.952</u>**	0.280	**<u>0.091</u>**
Original network	-	0.944	0.146	0.039

What do these numbers really mean for the network?

Filter bubbles

- We analyze the potential of weak ties on reducing filter bubbles



(Pariser 2011)

Diffusion experiment

- **Hypothesis**

> The more structurally diverse the recommendation is, the more diverse and novel the information flow through the network will be.

- **Experiment on interaction networks**
  1. Start with a baseline: Implicit MF / BM25
  2. Apply gradual rerankers for optimizing a metric
  3. Extend the network with top k recommended links
  4. Run propagation of (real) tweets through the network
  5. Measure diffusion properties (novelty & diversity)

Diffusion simulation

Simulate real information diffusion:

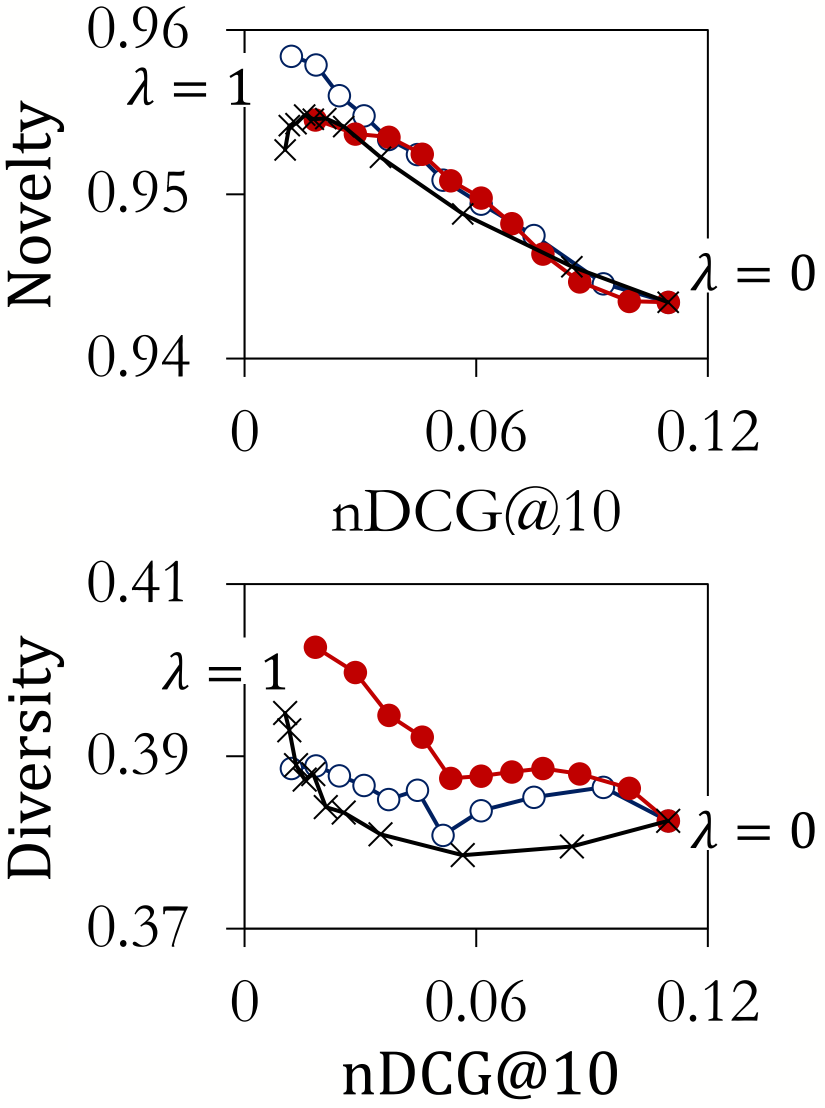
- Users spread their own tweets to their followers
- Users read everything that appears in their timeline
- Users retweet every tweet that they retweeted in real life

Evaluation: Information novelty and diversity

Measured in terms of hashtags (as topical information)

- **Novelty:**
  - Proportion of the hashtags unknown to the users.
  - **Known hashtags:** hashtags in their original tweets.
- **Diversity**
  - How evenly are hashtags propagated over the population
  - Complement of the Gini index

Results



Graph: Twitter 200-tweets interactions
Baseline: BM25

- Community Gini
- Modularity
- Clustering coefficient

Enhancing weak ties has positive effects in the novelty and diversity of the information flow

Conclusions

- **Accuracy is a partial perspective**

- **We propose evaluation perspectives beyond accuracy**
  - Global network effects beyond (averaged) isolated user gains
  - New metrics elaborating on weak ties

- **Enhancing the number of weak ties improves novelty & diversity of the information arriving to the users**

Conclusions

- **Accuracy is a partial perspective**

- **We propose evaluation perspectives beyond accuracy**
  - Global network effects beyond (averaged) isolated user gains
  - New metrics elaborating on weak ties

- **Enhancing the number of weak ties improves novelty & diversity of the information arriving to the users**

Conclusions

University of Glasgow

Summary

- **We have explored two problems in contact recommendation**
  - Definition of new models by adapting them from IR
  - Definition of new metrics for evaluating contact recommendation
- **IR models represent effective approaches for the task**
- **Enhancing the number of weak ties improves novelty & diversity of the information arriving to the users**
- **Do you want to try this?**

RELISON

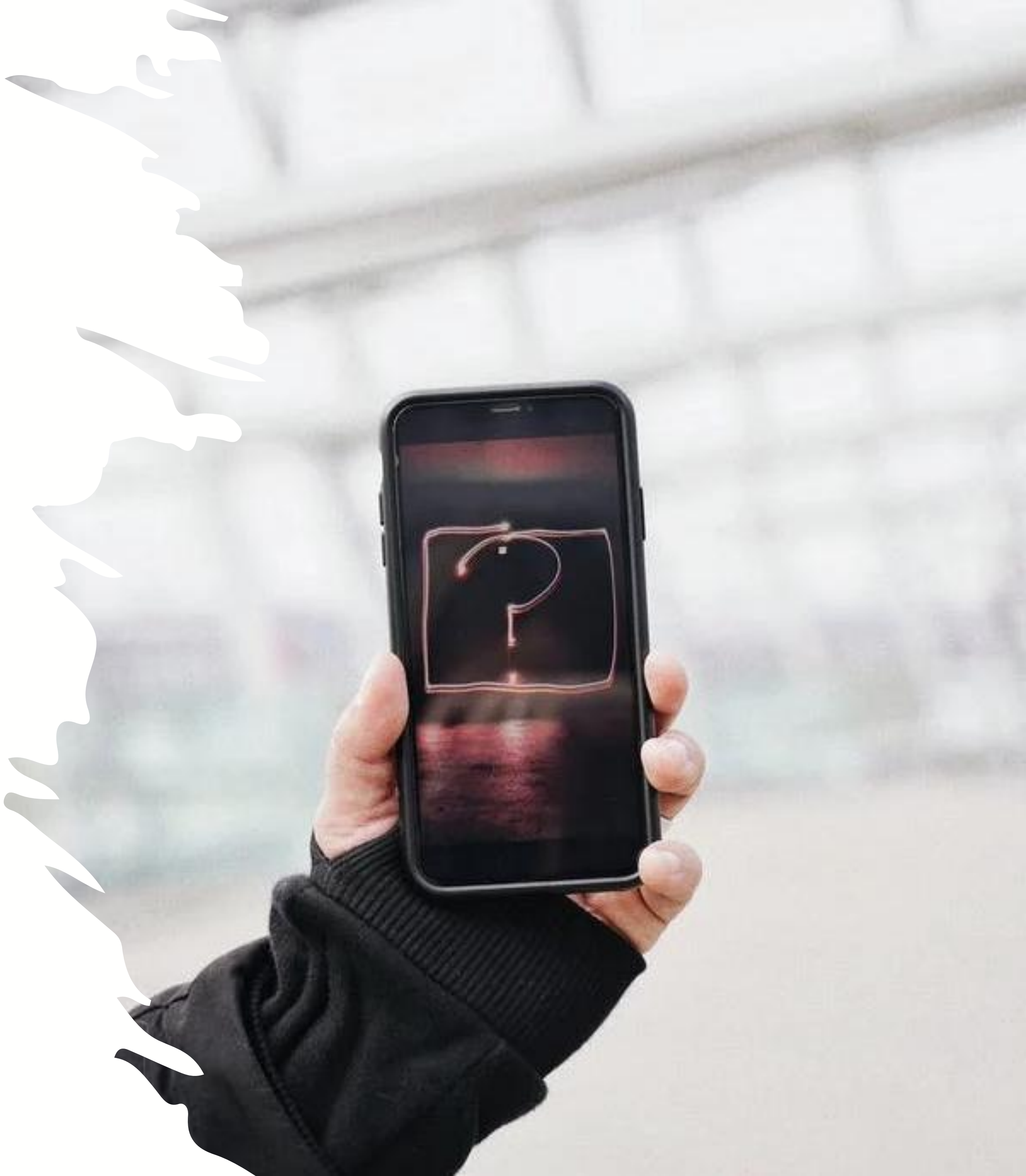
REcommending **LI**nks in **SO**cial **N**etworks

Questions?

Dr. Javier Sanz-Cruzado

Financial Recommendation Systems

javier.sanz-cruzadopuig@glasgow.ac.uk



Want to know more?

J. Sanz-Cruzado. **Contact recommendation in social networks: algorithmic models, diversity and network evolution**. 2021. PhD thesis. [Link]

Algorithmic models:

- J. Sanz-Cruzado, P. Castells, C. Macdonald, I. Ounis. **Effective Contact Recommendation in Social Networks by Adaptation of Information Retrieval Models.** *Information Processing & Management* , 57 (5), 102285, September 2020.
- J. Sanz-Cruzado, C. Macdonald, I. Ounis, P. Castells. **Axiomatic Analysis of Contact Recommendation Methods in Social Networks: An IR Perspective.** 42nd European Conference on Information Retrieval (ECIR 2020). Online, April 2020, pp. 157-190.
- J. Sanz-Cruzado, P. Castells. **Information Retrieval Models for Contact Recommendation in Social Networks**. 41st European Conference on Information Retrieval (ECIR 2019). Cologne, Germany, April 2019, pp. 148-163.

Want to know more? (II)

Network diversity:

- J. Sanz-Cruzado, P. Castells. **Enhancing Structural Diversity in Social Networks by Recommending Weak Ties.** *12th ACM Conference on Recommender Systems (RecSys 2018)*, Vancouver, Canada, October 2018, pp. 233-241.
- J. Sanz-Cruzado, P. Castells. **Beyond Accuracy in Link Prediction.** *3rd Workshop on Social Media for Personalization and Search (SoMePeAS 2019)* co-located with *41st European Conference on Information Retrieval (ECIR 2019)*. Cologne, Germany, April 2019, pp. 79-94.
- J. Sanz-Cruzado, S.M. Pepa, P. Castells. **Structural Novelty and Diversity in Link Prediction.** *9th International Workshop on Modeling Social Media (MSM 2018)* co-located with *The Web Conference 2018 (WWW 2018)*. Companion of The Web Conference 2018 . Lyon, France, April 2018, pp. 1347-1351.

References

- Adomavicius, G., Tuzhilin, A. **Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions.** *IEEE TKDE* 17(6), 2005.
- Bellogín, A., Wang, J., Castells, P. **Bridging memory-based collaborative filtering and text retrieval.** *Information Retrieval* 16(6), pp. 697-724, 2013.
- Burges, C. **From RankNet to LambdaRank to LambdaMART: an overview**. *Microsoft Technical Report*, 2010.
- De Meo, P., Ferrara, E., Fiumara, G., Provetti, A. **On Facebook, Most Ties are Weak.** *Communications of the ACM* 57(11), pp. 78-84, 2012.
- Ganjisaffar, Y., Caruana, R., Lopes, C. **Bagging gradient-boosted trees for high precision, low variance ranking models.** *SIGIR* 2011, Beijing, China, pp. 85-94, 2011.
- Granovetter, M. **The Strength of Weak Ties**. *American Journal of Sociology* 78(6), pp. 1360-1380, 1973.
- Guy, I. **People recommendation on social media.** *Social Information Access*, pp. 570-623, 2018.
- Hannon, J., Bennet, M., Smyth, B. **Recommending Twitter users to follow using content and collaborative filtering approaches.** *ACM RecSys 2010*, Barcelona, Spain, pp. 199-206, 2010.

References (II)

- Liu, T. **Learning to rank for information retrieval.** *Foundations and Trends in Information Retrieval* 3(3), pp. 225-331, 2007.
- Ning, X., Desrosiers, C., Karypis, G. **A comprehensive survey of neighborhood-based recommendation methods.** *Recommender Systems Handbook*, 2nd edition, pp. 37-76, 2015.
- Parapar, J., Bellogín, A., Castells, P., Barreiro, A. **Relevance-based language modelling for recommender systems.** *Information Processing & Management* 49, pp. 966-980, 2013.
- Pariser, E. **The Filter Bubble: How the new personalized web is changing what we read and how we think.** 2011.
- Sanz-Cruzado, J., Castells, P. **Contact Recommendations in Social Networks.** In S. Berkovsky, I. Cantador, D. Tikk (Eds): *Collaborative Recommendations: Algorithms, Practical Challenges and Applications*, World Scientific Publishing, pp. 519-570, 2018.
- Valcarce, D., Parapar, J., Barreiro, A. **Axiomatic analysis of language modelling of recommender systems**. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 25, pp. 113-127, 2017.
- Wang, J., Robertson, S., de Vries, A.P., Reinders, M.J.T. **Probabilistic relevance ranking for collaborative filtering**. *ACM TOIS*, 26, article 16. 2008.

An example: BM25

Text IR:

$$f_q(d) = \sum_{t \in d \cap q} \frac{(k+1)\,\mathrm{freq}(t,d)}{k\left(1 - b + \frac{b|d|}{\mathrm{avg}_{d'}|d'|}\right) + \mathrm{freq}(t,d)} \mathrm{RSJ}(t)$$

$$\mathrm{RSJ}(w) = \log \frac{|D| - |D_t| - 0.5}{|D_t| - 0.5}$$

d: document → $\Gamma(v)$: candidate user

q: query → $\Gamma(u)$: target user

$t \in q \cap d$: term → $t \in \Gamma(u) \cap \Gamma(v)$: neighbor user

D: all documents → $\mathcal{U}$: all users

D_t: documents containing t → $\Gamma(t)$: v containing t in $\Gamma(v)$

$\mathrm{freq}(t,d)$: frequency of $t \in d$ → $w(t,v)$: edge weight

$|d|$: document d length → $\mathrm{len}(v) = \sum_{x \in \Gamma(v)} w(x,v)$

An example: BM25

Text IR:

$$f_q(d) = \sum_{t \in d \cap q} \frac{(k+1)\,\mathrm{freq}(t,d)}{k\left(1 - b + \frac{b|d|}{\mathrm{avg}_{d'}|d'|}\right) + \mathrm{freq}(t,d)} \mathrm{RSJ}(t)$$

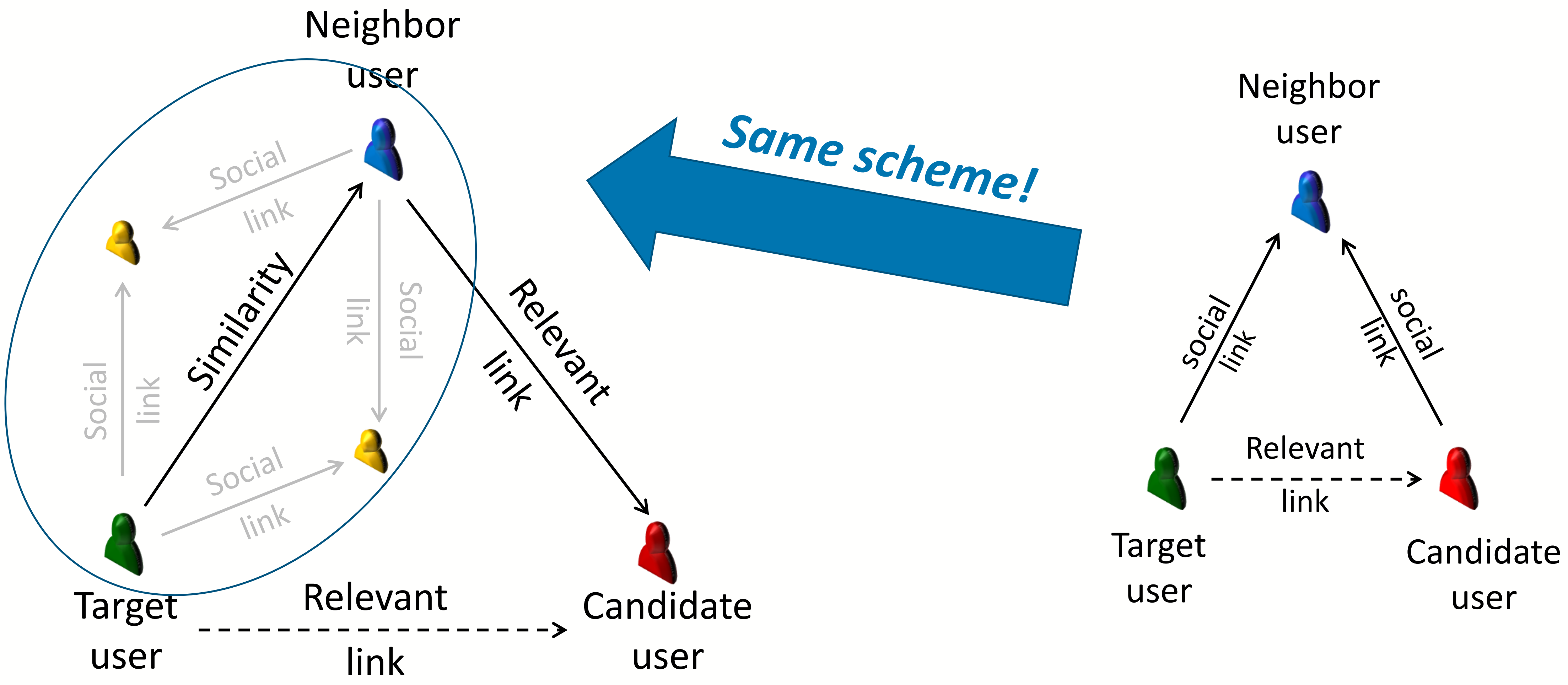
$$\mathrm{RSJ}(w) = \log \frac{|D| - |D_t| - 0.5}{|D_t| - 0.5}$$

Contact recommendation:

$$f_u(v) = \sum_{t \in \Gamma(u) \cap \Gamma(v)} \frac{(k+1) w(t,v)}{k\left(1 - b + \frac{b \cdot \mathrm{len}(v)}{\mathrm{avg}_{v'}\left(\mathrm{len}(v')\right)}\right) + w(t,v)} \mathrm{RSJ}(t)$$

$$\mathrm{RSJ}(t) = \log \frac{|\mathcal{U}| - |\Gamma(t)| + 0.5}{|\Gamma(t)| + 0.5}$$

K Nearest Neighbors



(Ning et al. 2015)