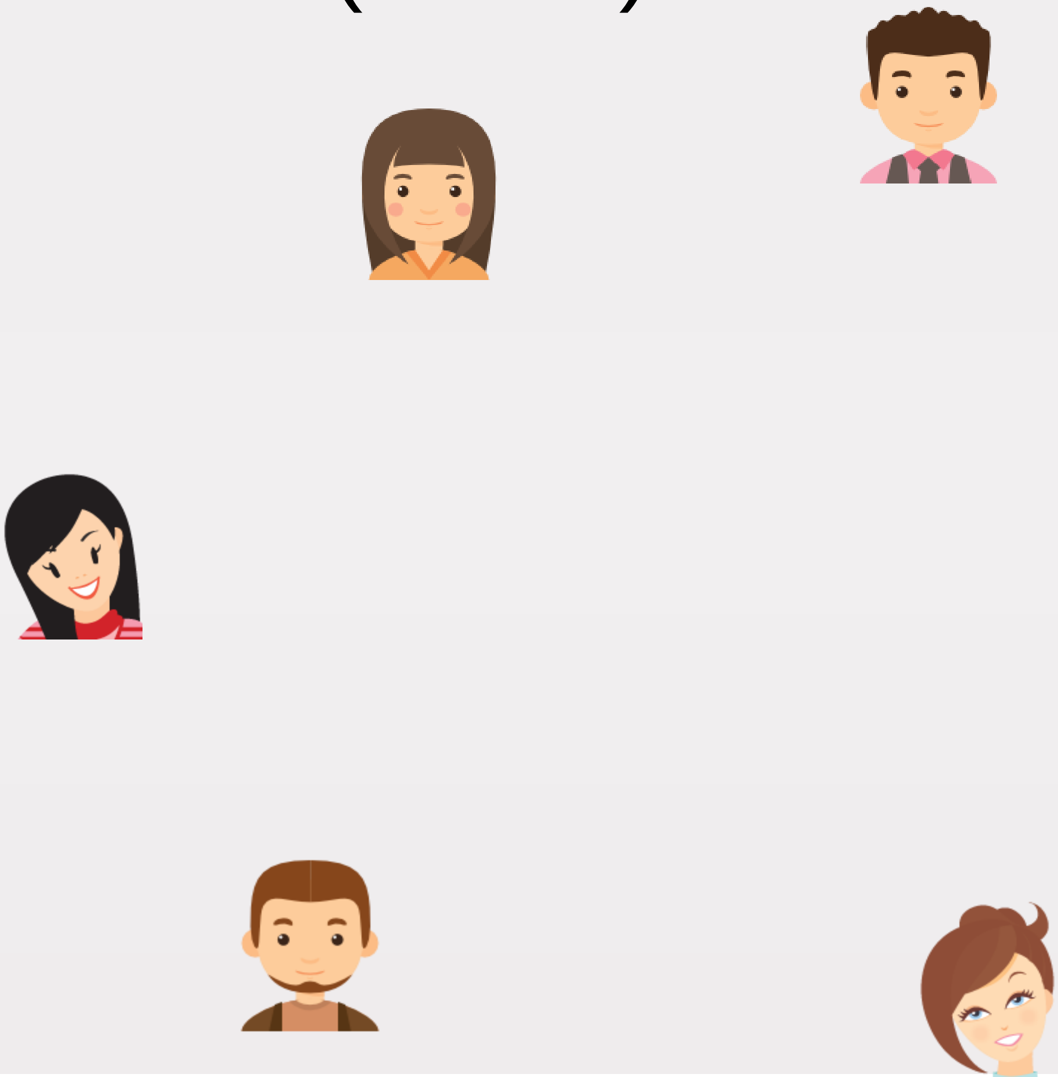
Graph-based Analytics for Decentralized Online Social Networks

Amira Soliman

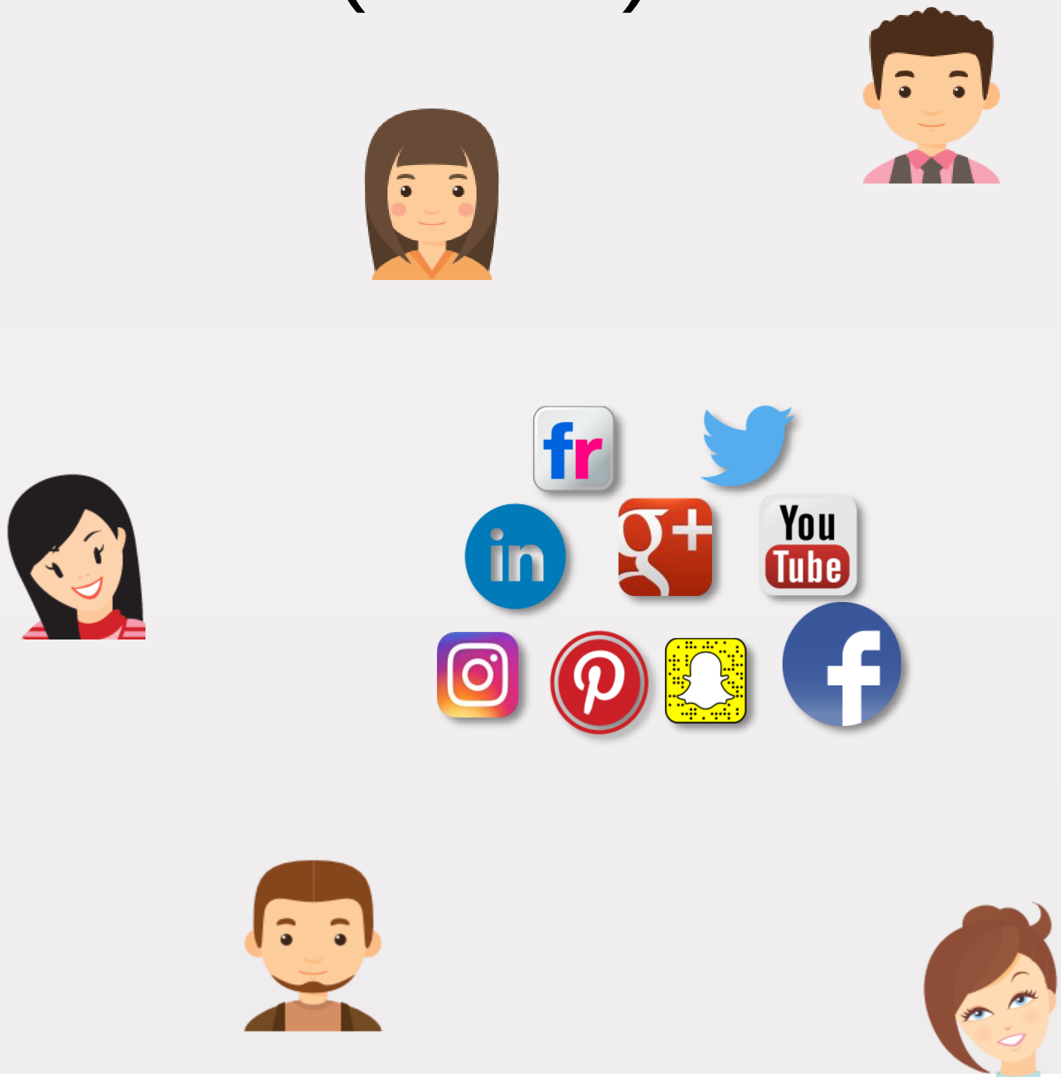
Researcher @ RISE SICS.

The 6th Swedish Workshop on Data Science (SweDS2018)
Umeå University, Umeå, Sweden.
November 20-21, 2018.

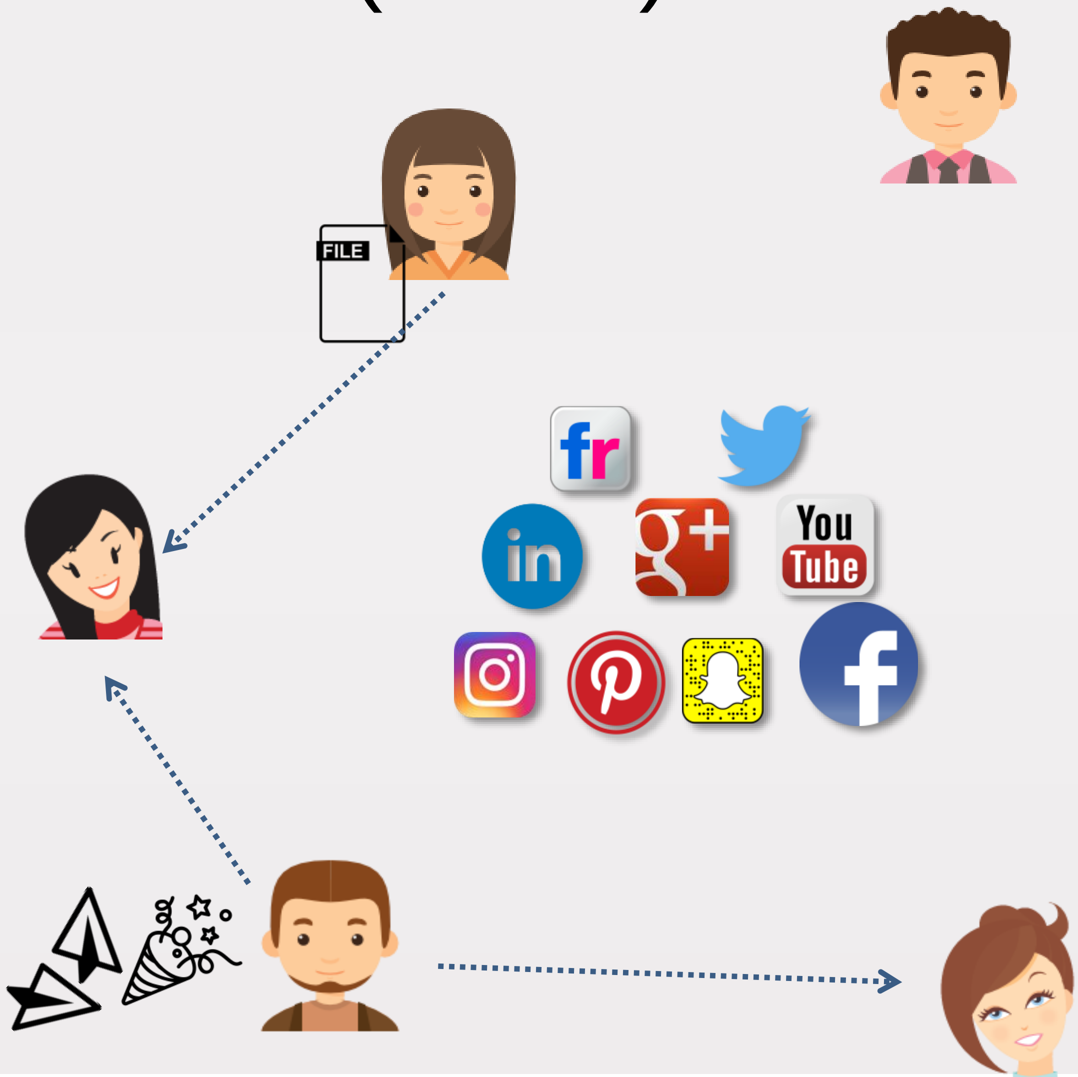
Interactions on top of Online Social Networks (OSNs)



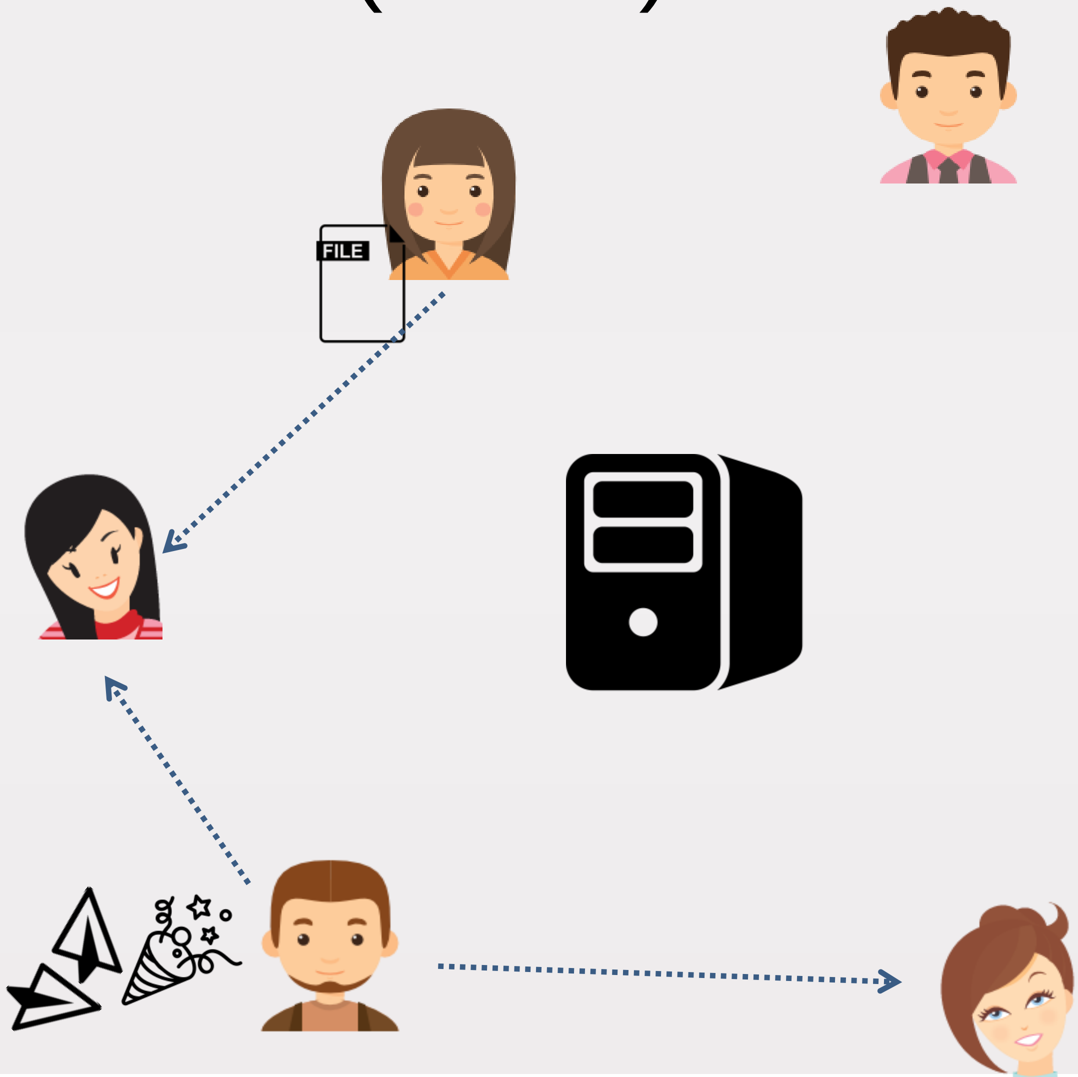
Interactions on top of Online Social Networks (OSNs)



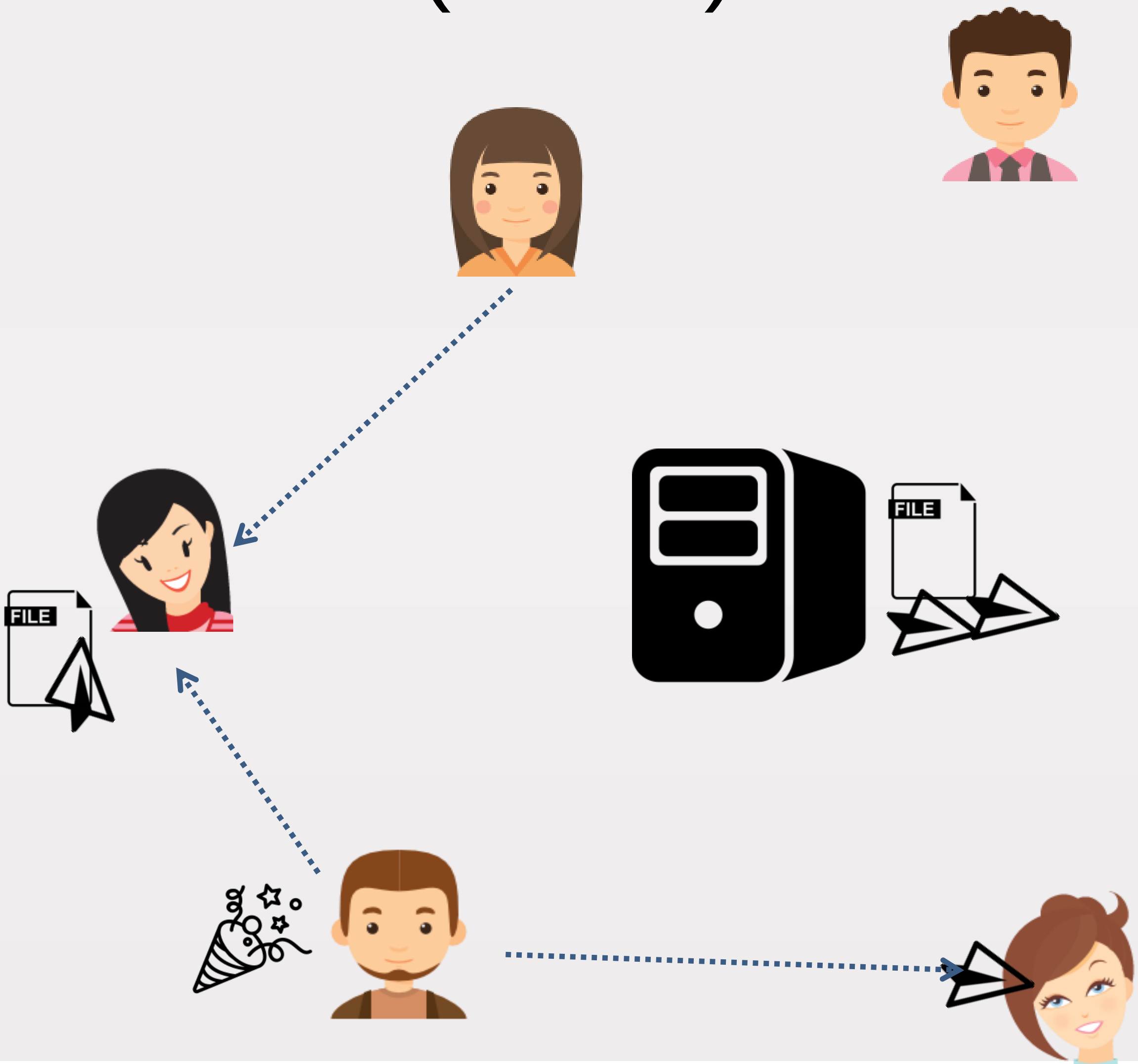
Interactions on top of Online Social Networks (OSNs)



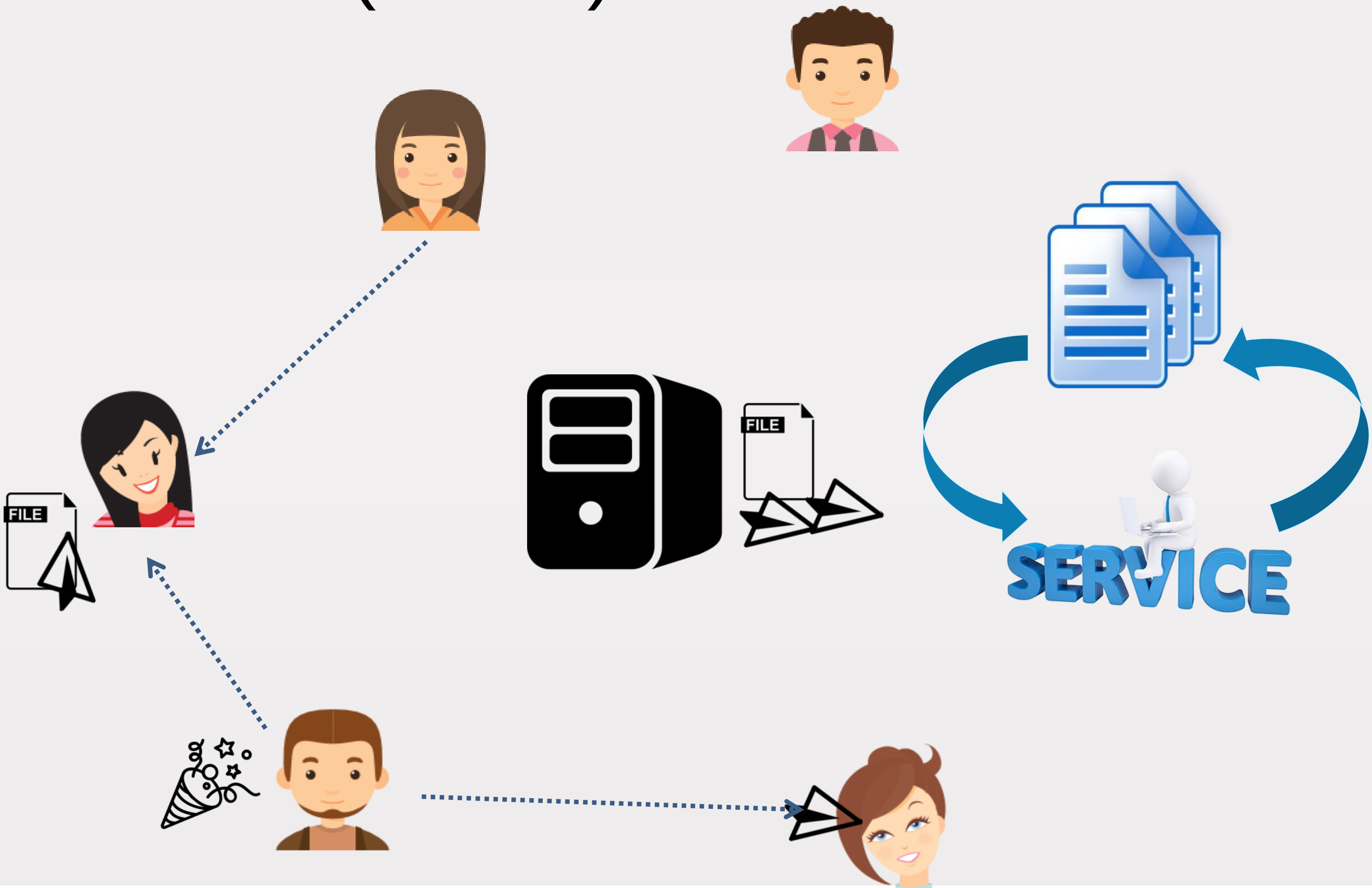
Interactions on top of Online Social Networks (OSNs)



Interactions on top of Online Social Networks (OSNs)



Interactions on top of Online Social Networks (OSNs)



Data Repurposing: we have no idea about how our data is being used!

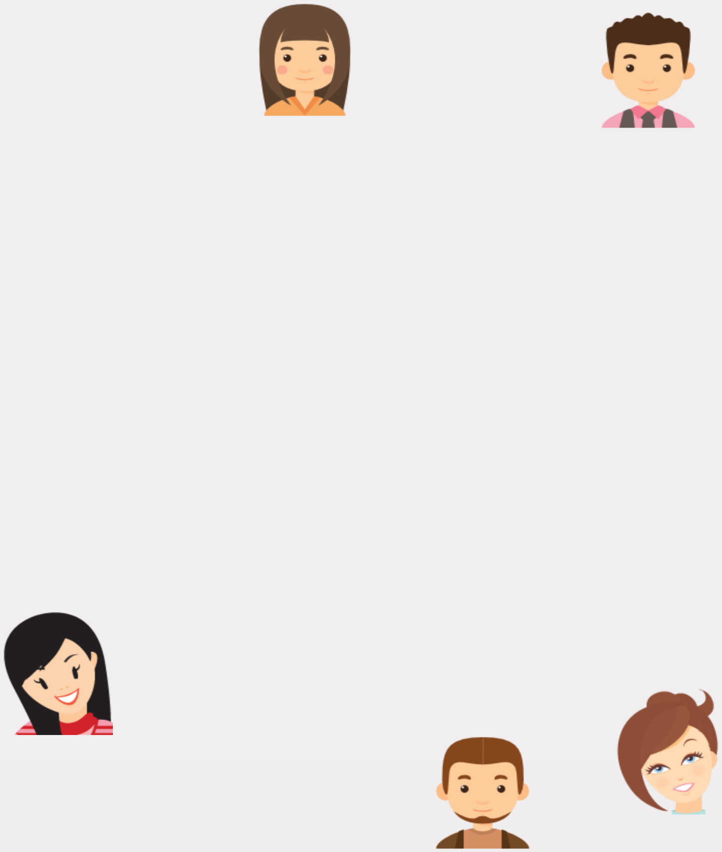
- Providers analyze what people share for public, yet they also scan private messages.
- Providers collect personal tastes and navigation from third party applications.
- AI can guess whether a person is gay or straight using profile picture.
- Tagging can violate the privacy of your friends.

Data Repurposing: we have no idea about how our data is being used!

- Providers analyze what people share for public, yet they also scan private messages.
- Providers collect personal [illegible]es and navigati[illegible] from third party applications.
- AI can guess whether [illegible] person is gay or straight [illegible]ng profile picture.
- Tagging can violate [illegible] privacy of your friends[illegible]

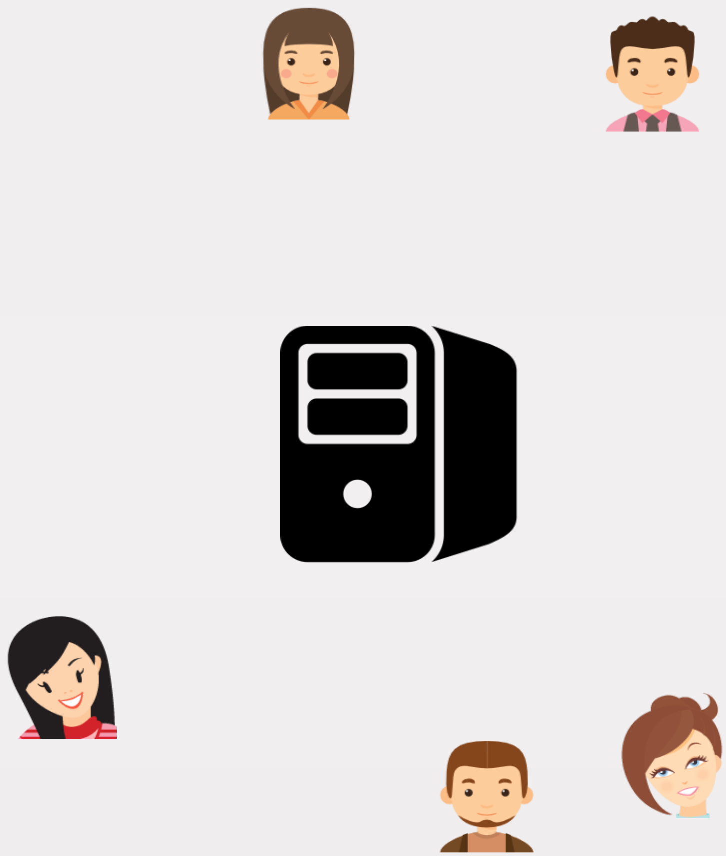
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.



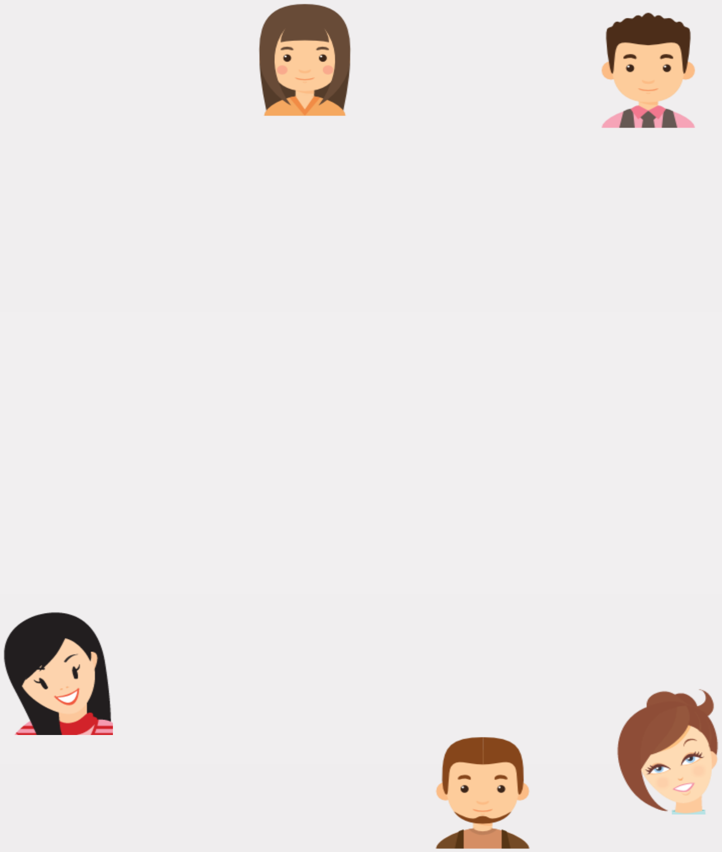
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.



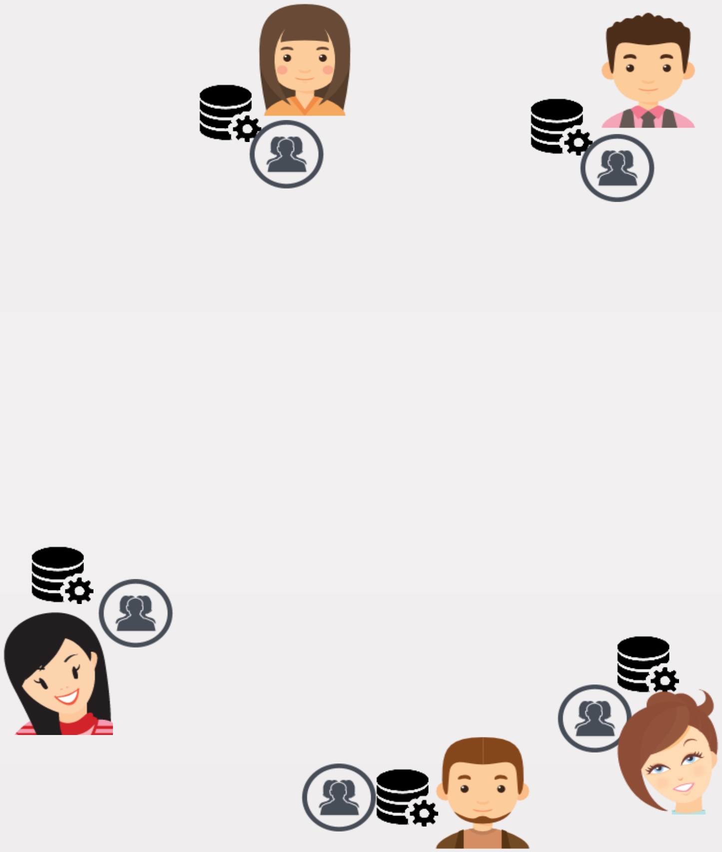
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.



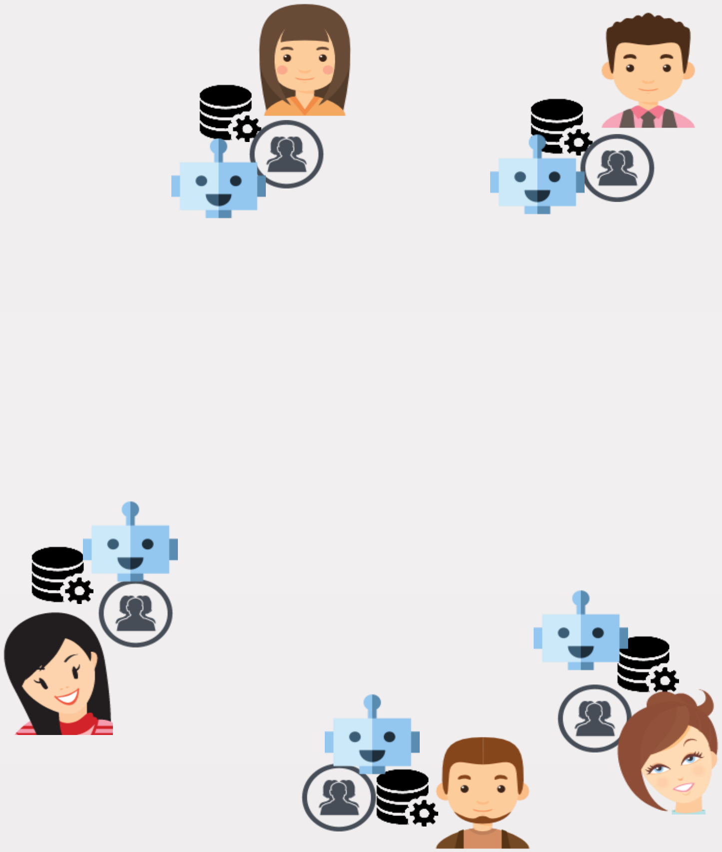
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.



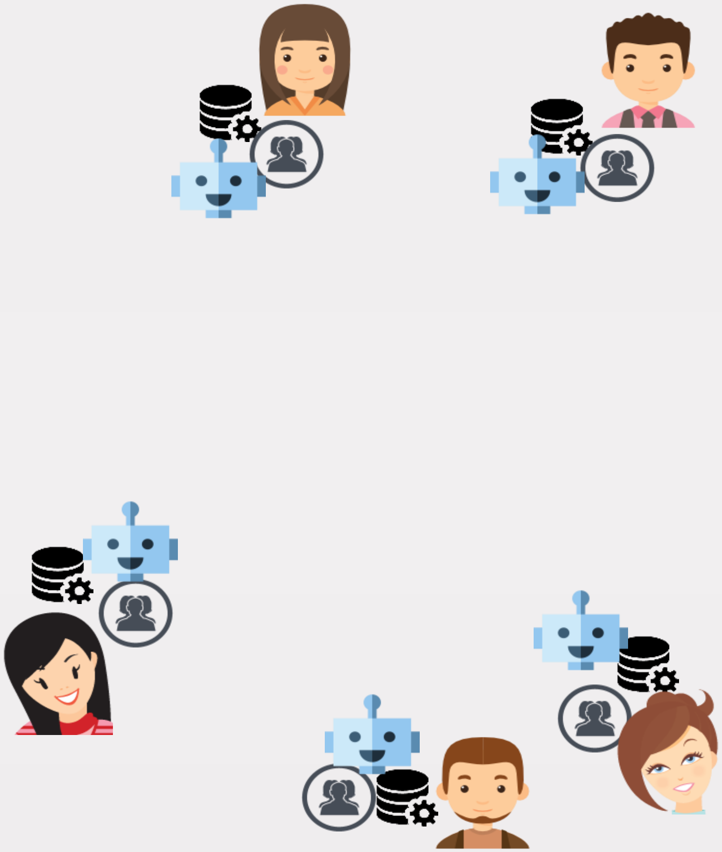
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.



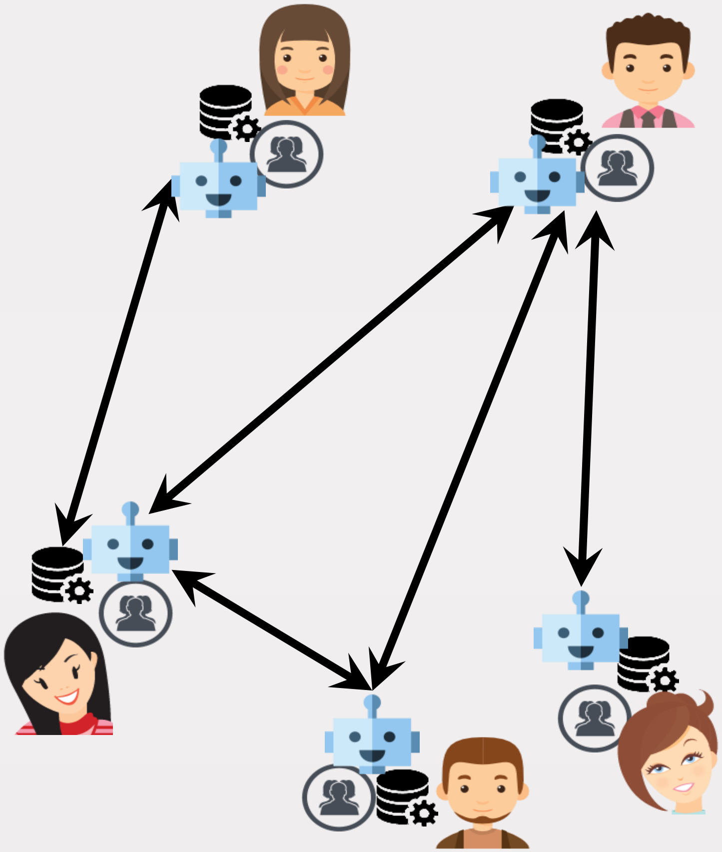
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.



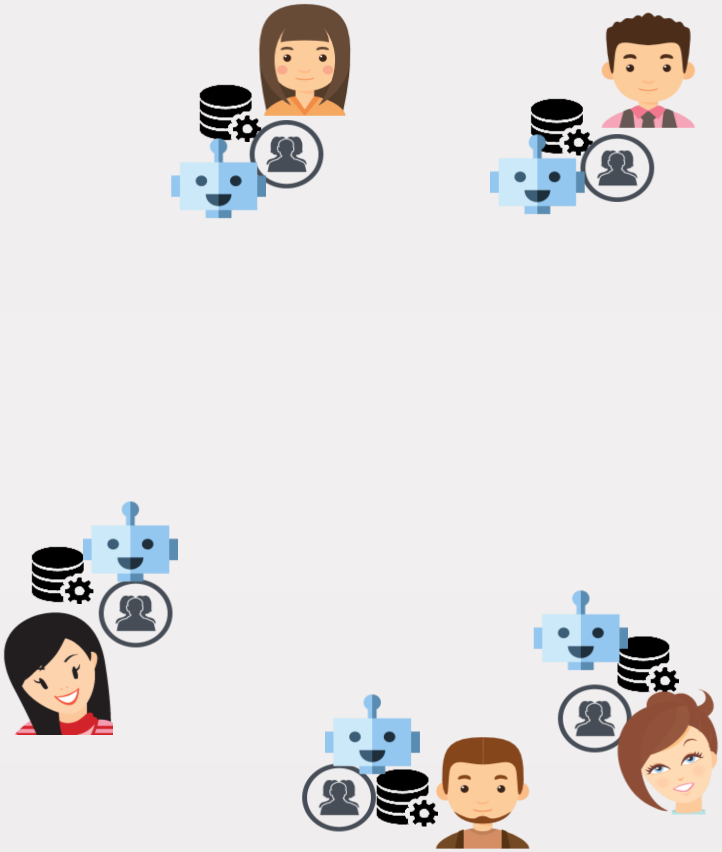
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.



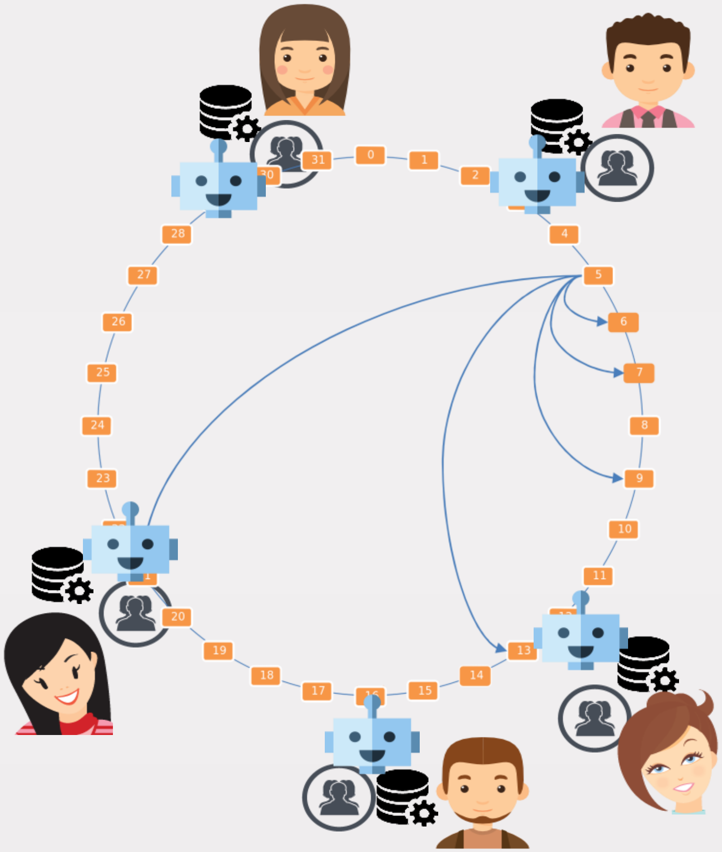
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.



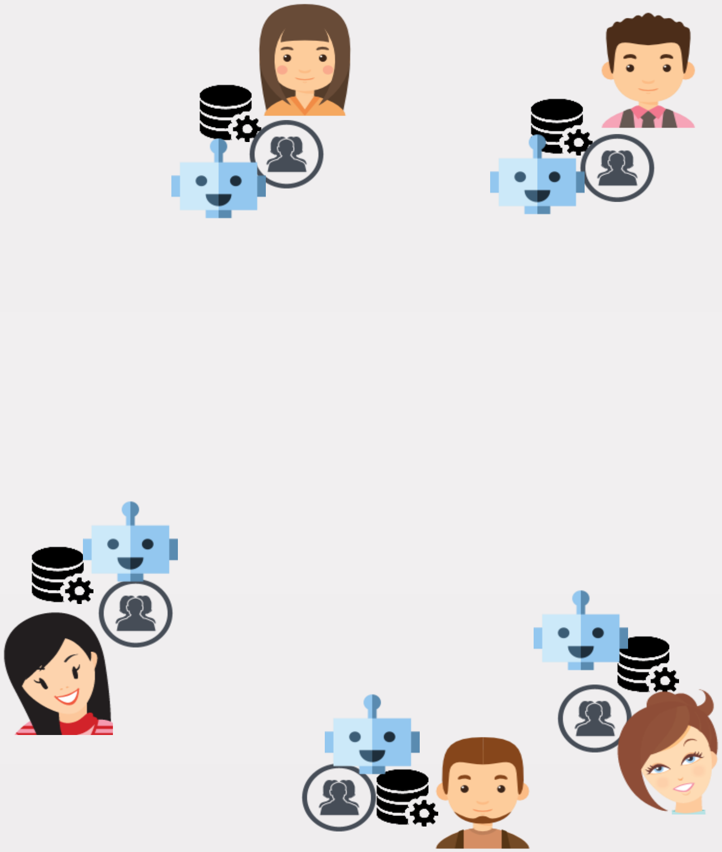
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.



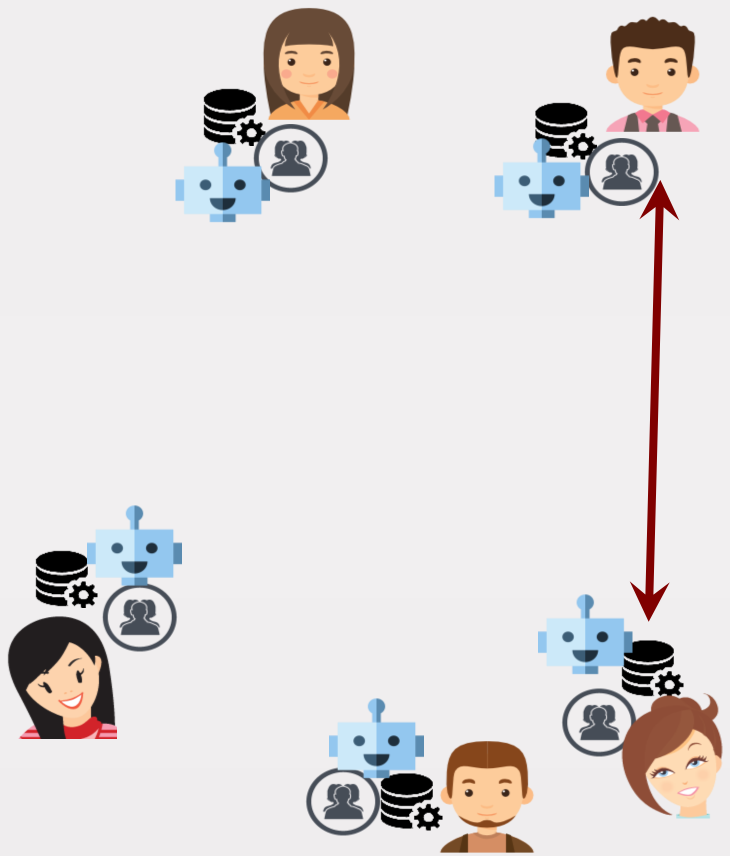
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.



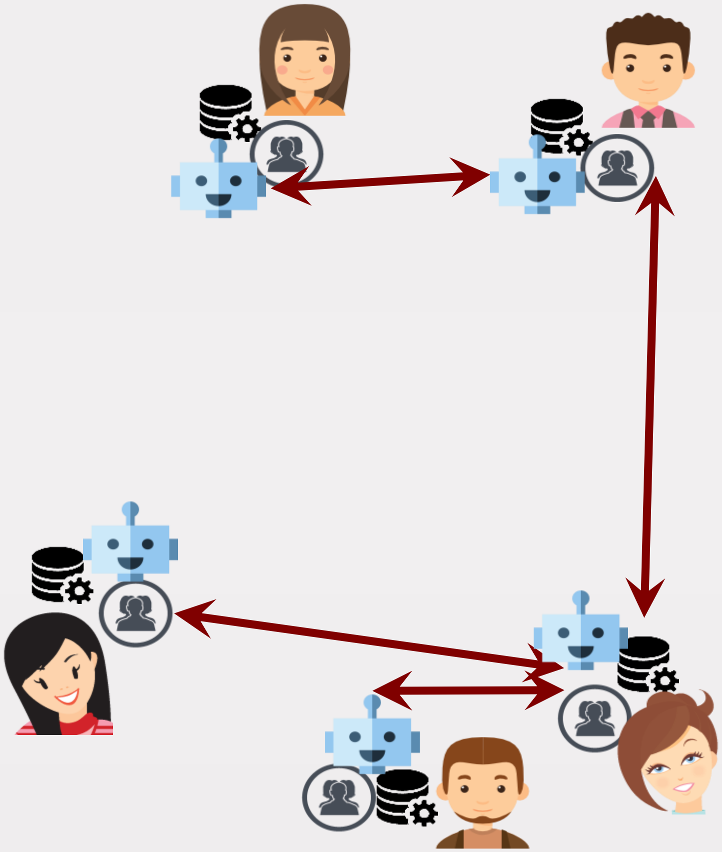
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.



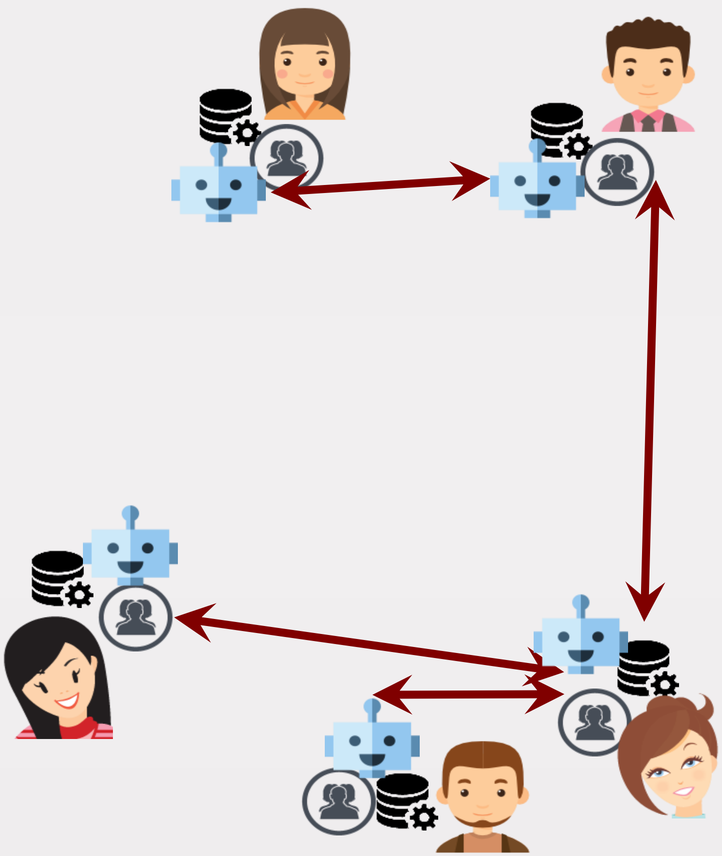
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.



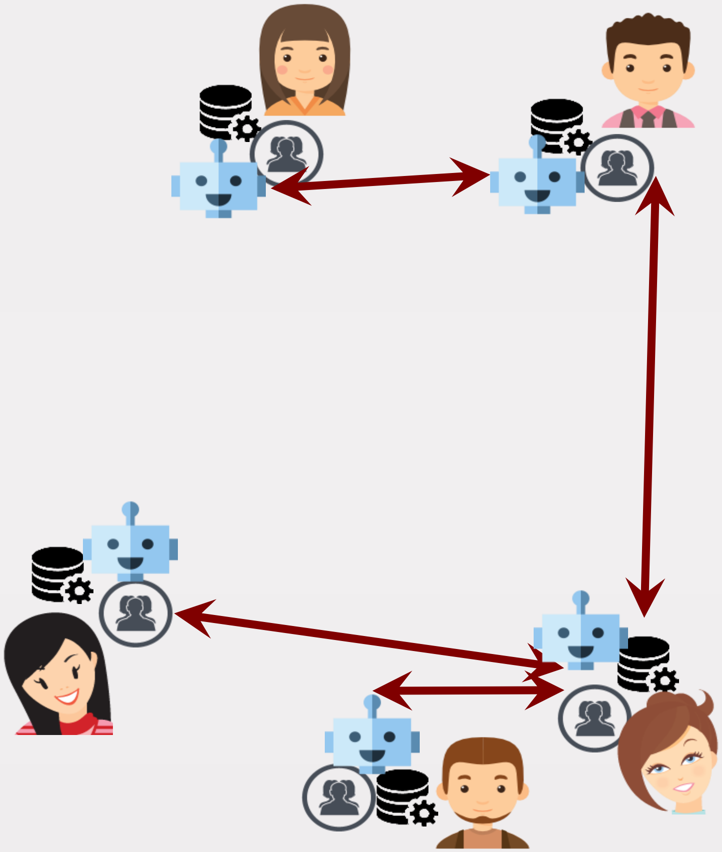
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.
- E.g., Diaspora active user rate:



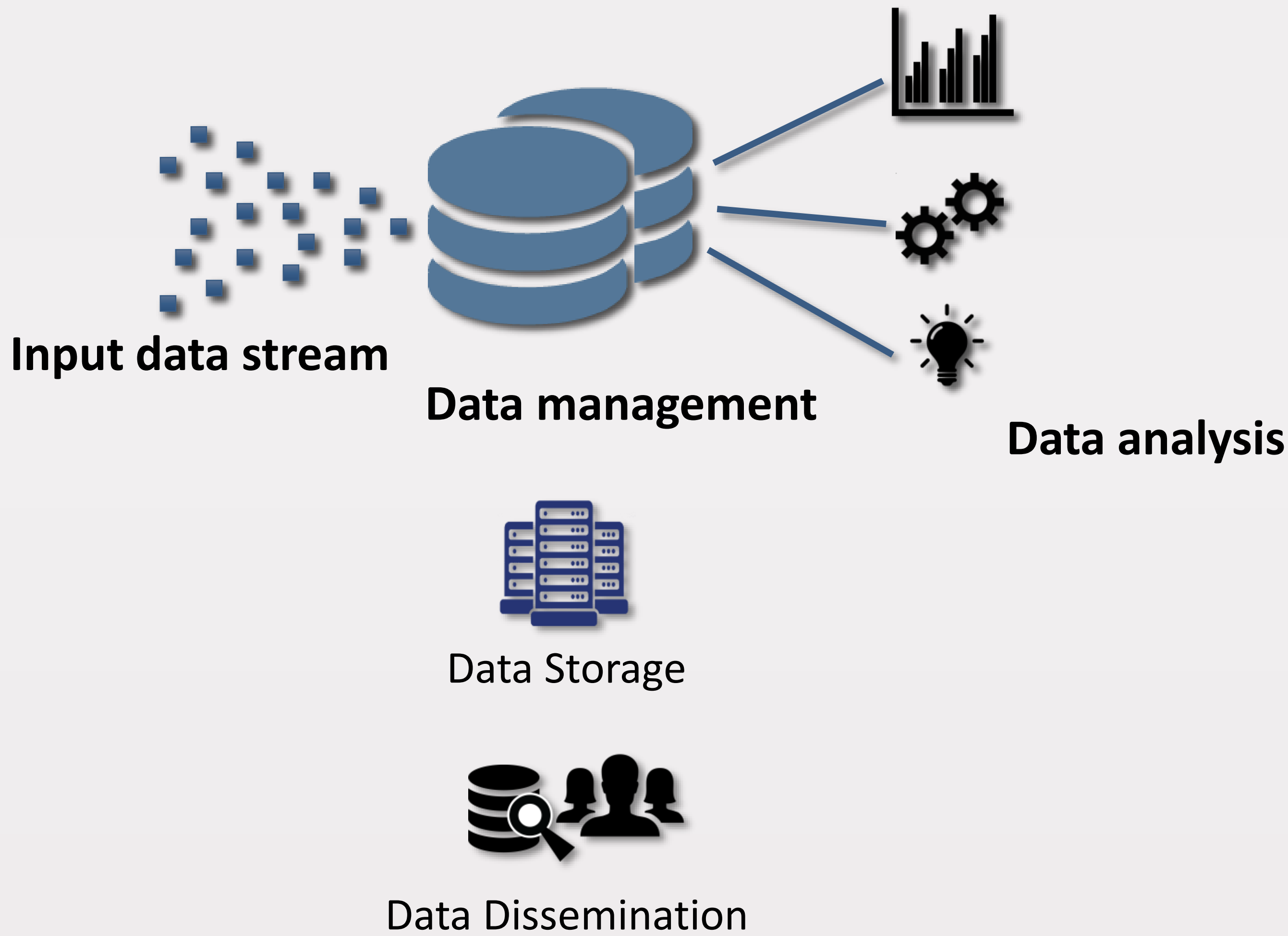
Decentralized OSNs – (DOSNs)

- Users have control over privacy, ownership, and dissemination of their data.
- DOSNs can be modeled as distributed systems using various topological overlays.
- E.g., Diaspora active user rate:
  - During 2014: 1M.
  - By end of 2015 till now: 660K.

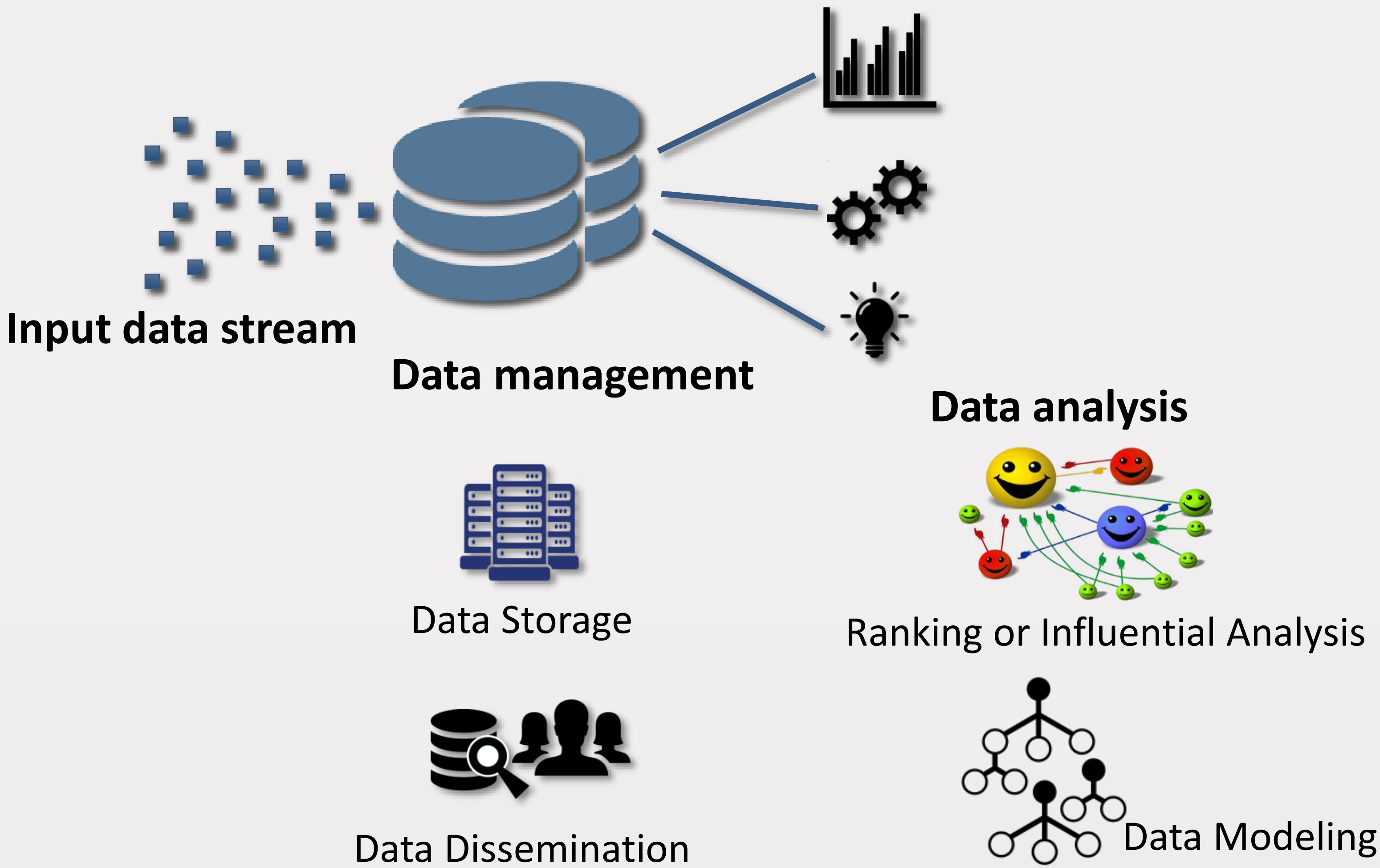


Service Categories for Social Networks

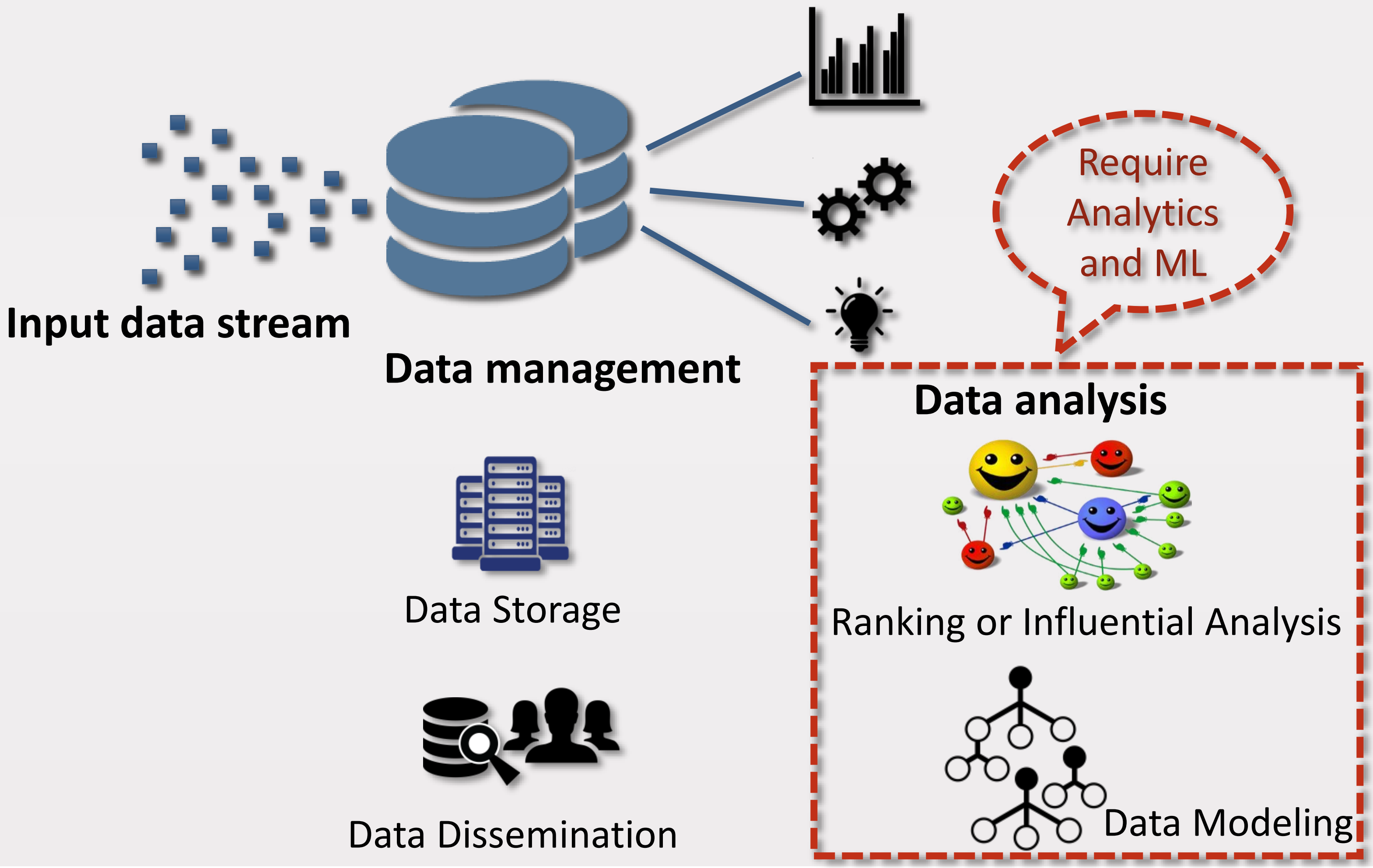
Service Categories for Social Networks



Service Categories for Social Networks

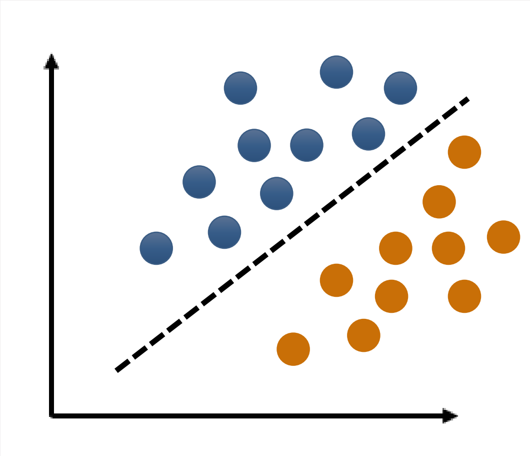


Service Categories for Social Networks



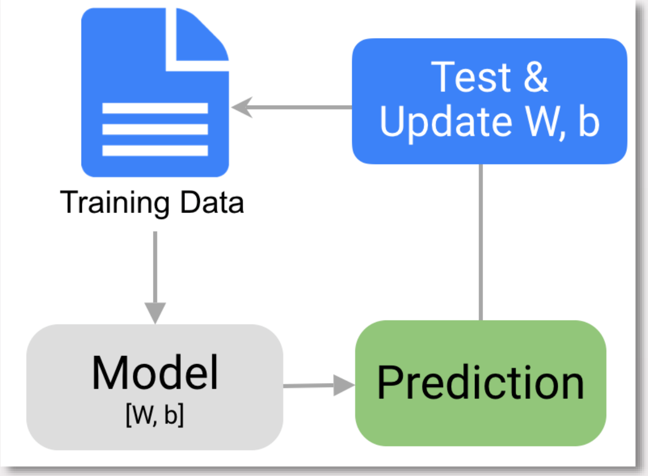
Machine Learning (ML) and DOSNs

- ML applications are iterative.



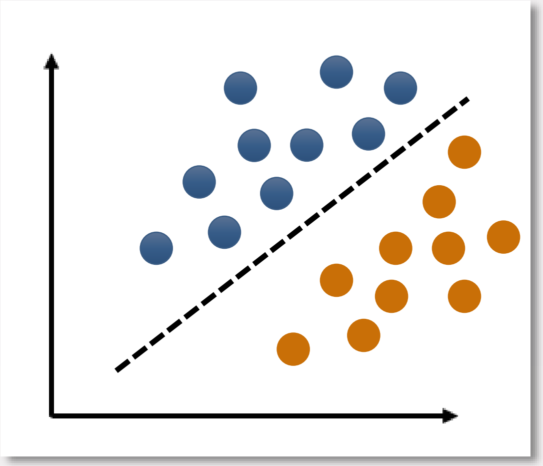
$$y = W * x + b$$

OUTPUT SLOPE INPUT Y-INTERCEPT



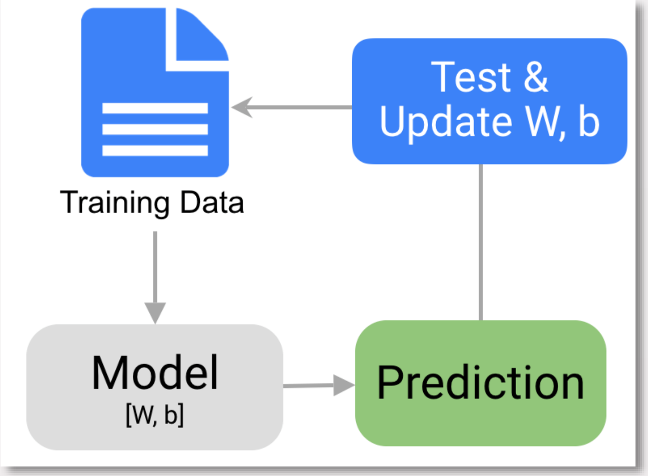
Machine Learning (ML) and DOSNs

- ML applications are iterative.
- DOSNs introduce new challenges.



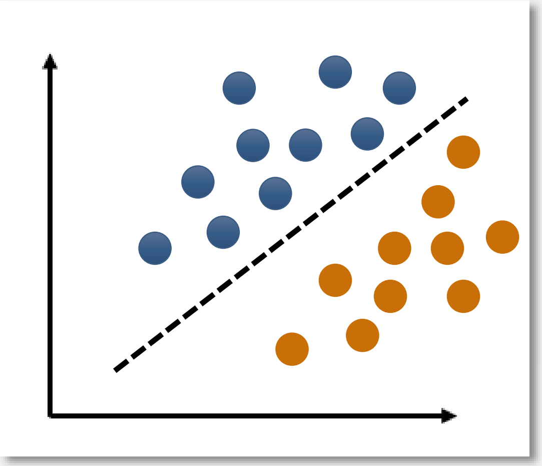
$$y = W * x + b$$

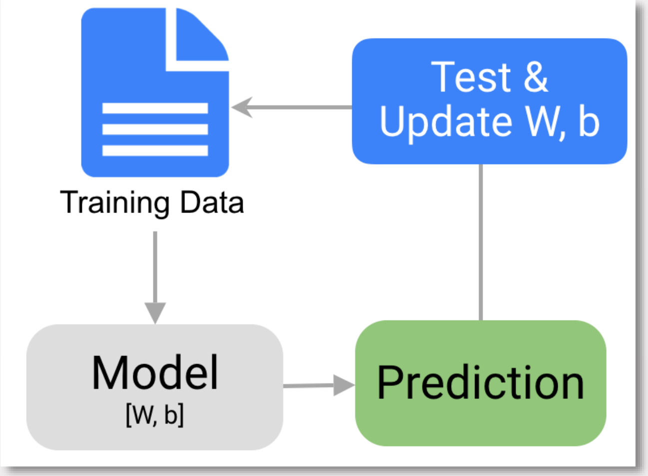
OUTPUT SLOPE INPUT Y-INTERCEPT

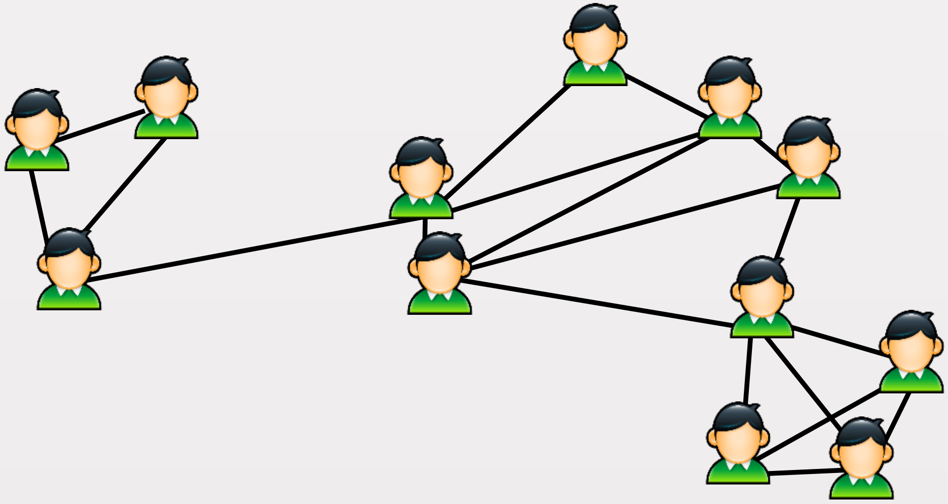


Machine Learning (ML) and DOSNs

- ML applications are iterative.
- DOSNs introduce new challenges.

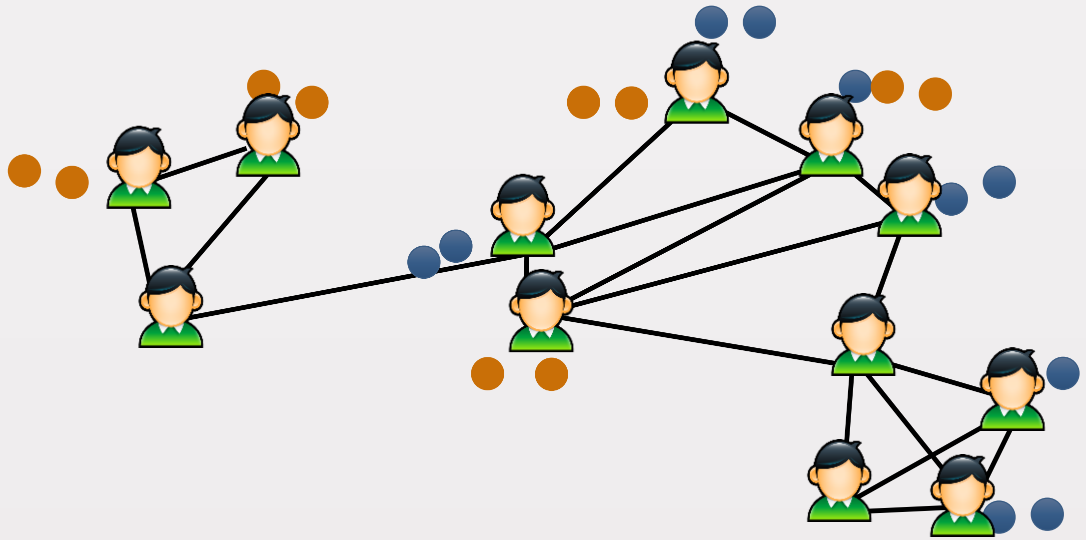


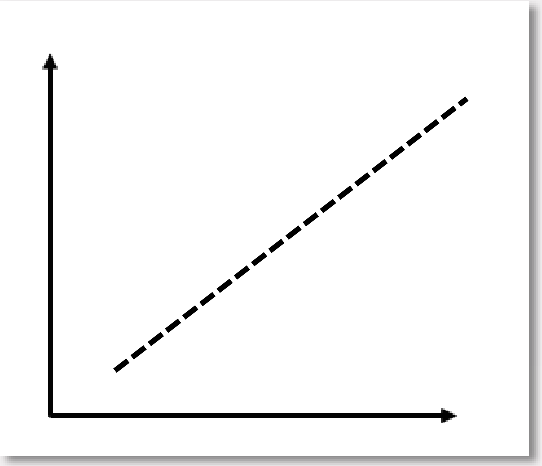




Machine Learning (ML) and DOSNs

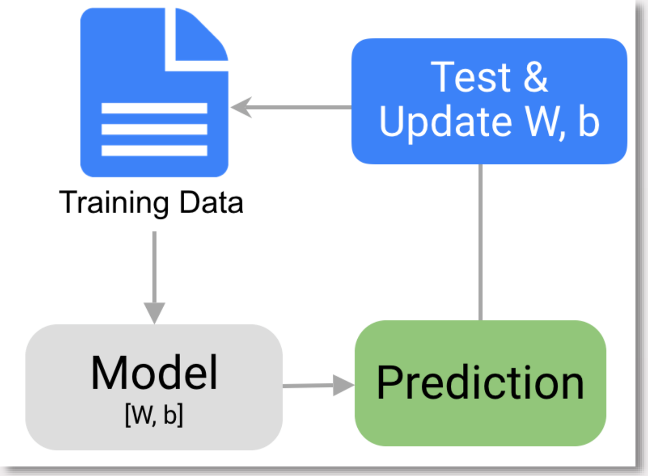
- ML applications are iterative.
- DOSNs introduce new challenges.





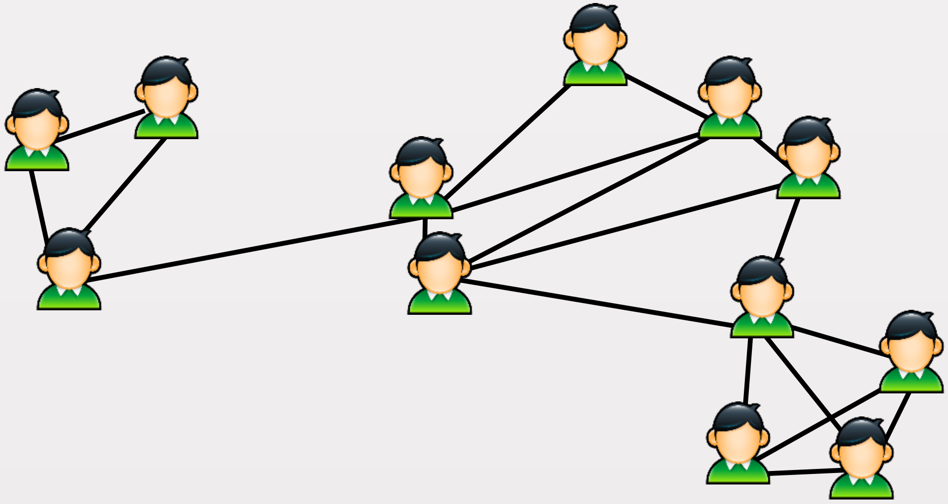
$$y = W * x + b$$

OUTPUT SLOPE INPUT Y-INTERCEPT



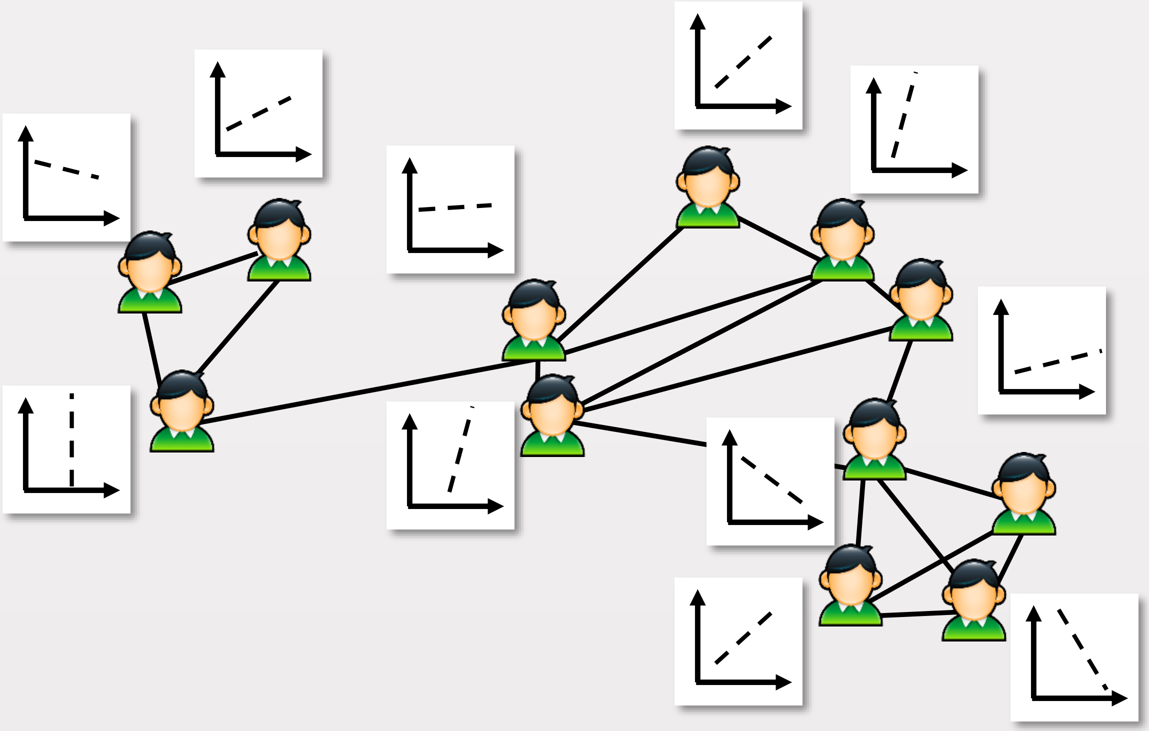
Machine Learning (ML) and DOSNs

- ML applications are iterative.
- DOSNs introduce new challenges.



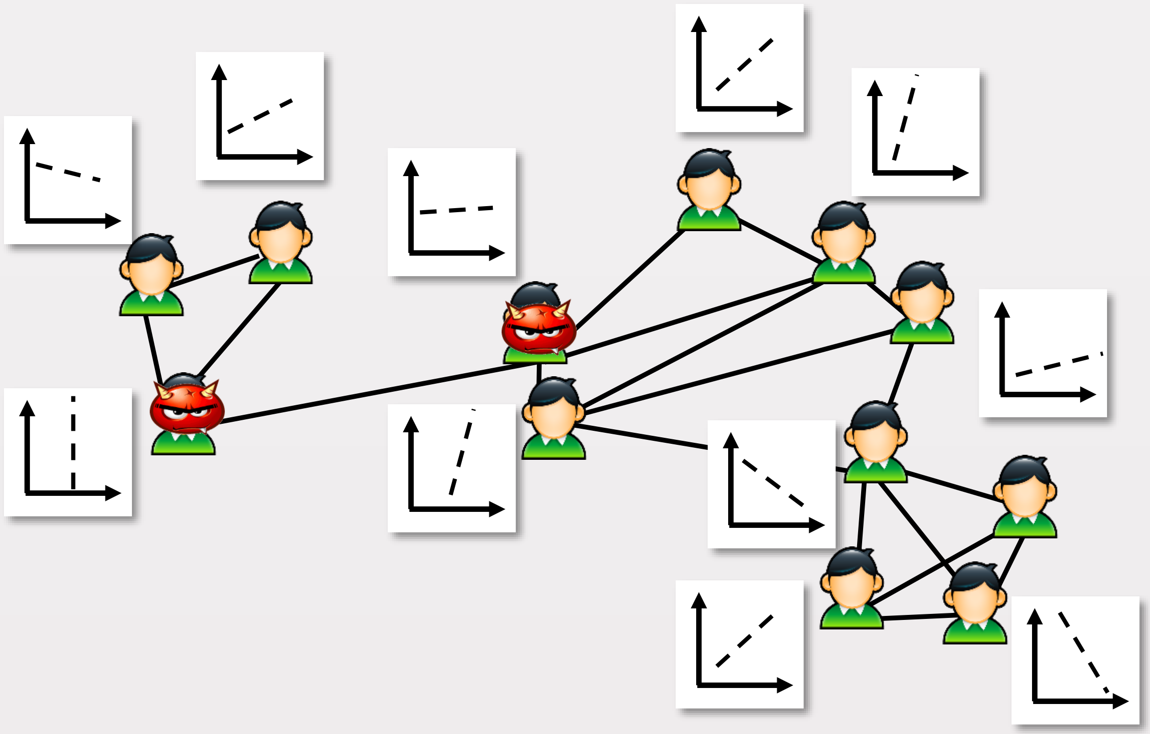
Machine Learning (ML) and DOSNs

- ML applications are iterative.
- DOSNs introduce new challenges.



Machine Learning (ML) and DOSNs

- ML applications are iterative.
- DOSNs introduce new challenges.

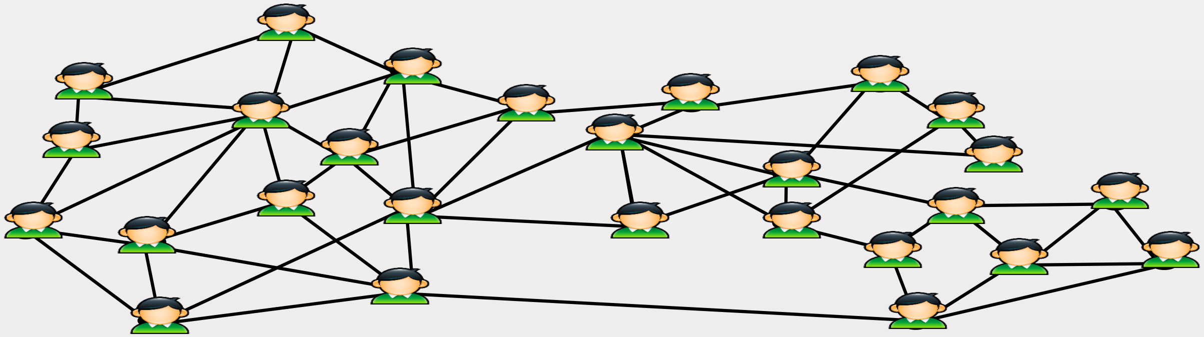


Machine Learning (ML) and DOSNs

- ML applications are iterative.
- DOSNs introduce new challenges.
  - Data is fully distributed.
  - Data movement is constrained.
  - Communication is limited.
  - Deviation from workflow protocol is possible.

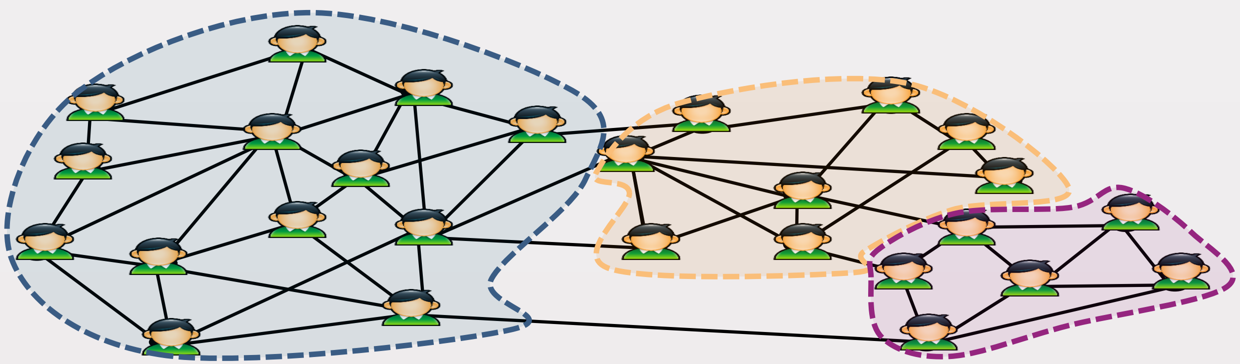
Community-aware Learning

- Homophily: from ancient Greek ὁμοῦ (homou, “together”) and Greek φιλία (philia, "friendship") is the tendency of individuals to associate and bond with similar others [Wikipedia].
- Integrating community structure with analytic tasks enhances data analytic insights and improves results (e.g., validating online identities and spam detection).



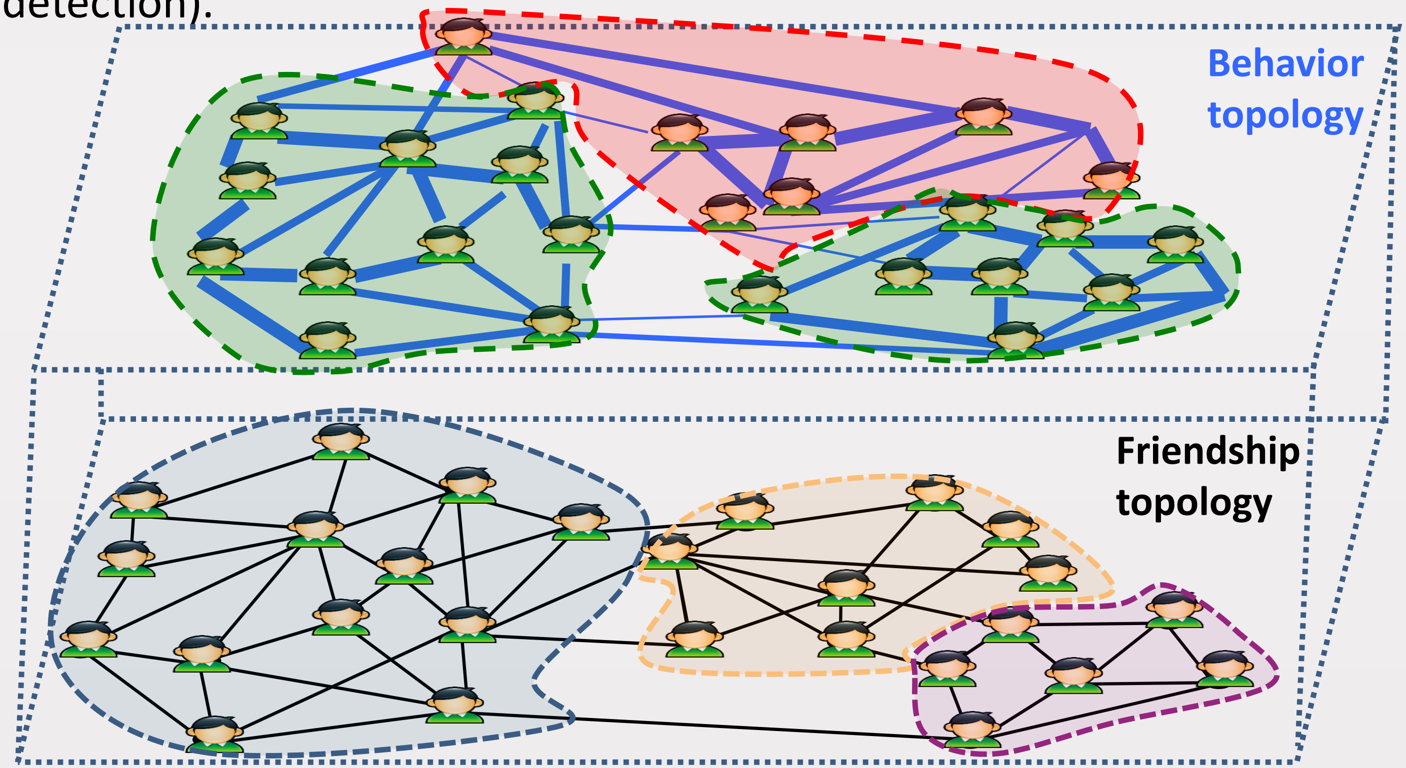
Community-aware Learning

- Homophily: from ancient Greek ὁμοῦ (homou, “together”) and Greek φιλία (philia, "friendship") is the tendency of individuals to associate and bond with similar others [Wikipedia].
- Integrating community structure with analytic tasks enhances data analytic insights and improves results (e.g., validating online identities and spam detection).



Community-aware Learning

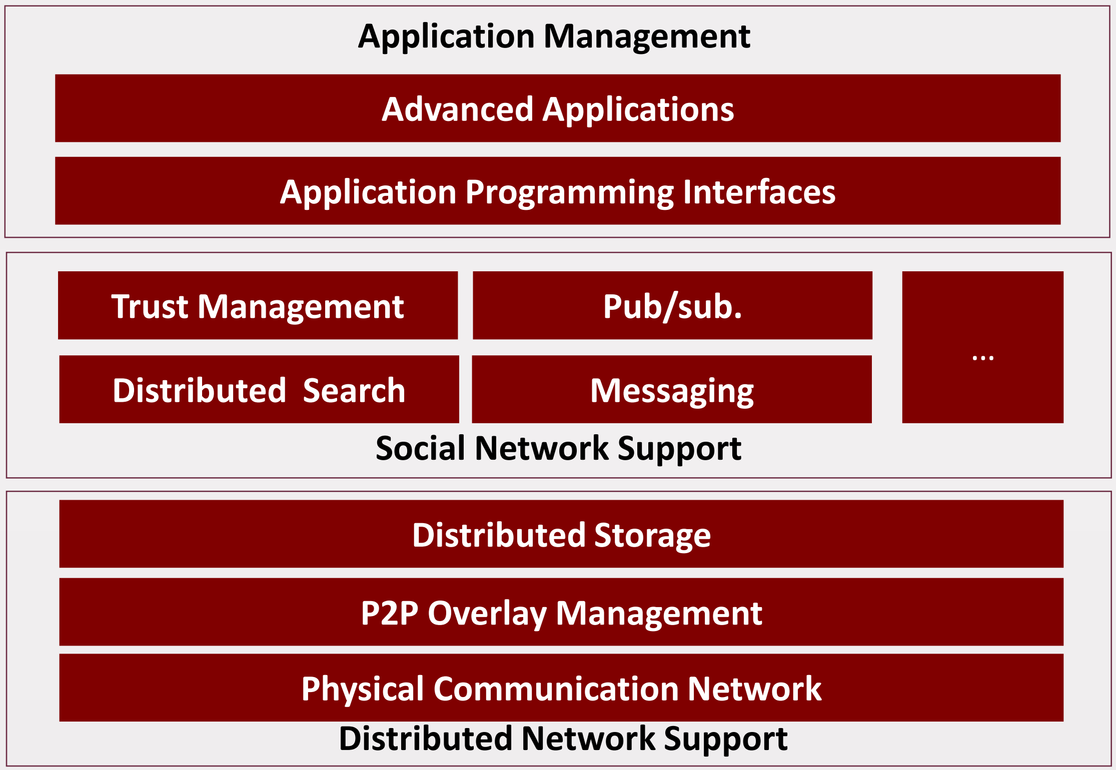
- Homophily: from ancient Greek ὁμοῦ (homou, “together”) and Greek φιλία (philia, "friendship") is the tendency of individuals to associate and bond with similar others [Wikipedia].
- Integrating community structure with analytic tasks enhances data analytic insights and improves results (e.g., validating online identities and spam detection).



Main Objectives

DOSN requires graph-based algorithms and methods that allow generation of efficient ML models using user local data in fully decentralized, iterative, massively parallel, and highly scalable manner, thereby ***enabling decentralized ML and analytic tasks on DOSNs, without violating privacy preservation constraints***.

The Stack of Services



A. Datta, S. Buchegger, L.-H. Vu, T. Strufe, and K. Rzadca, "Decentralized online social networks," in Handbook of Social Network Technologies and Applications . Springer, 2010, pp. 349–378.

The Stack of Services

Application Management

- Advanced Applications
- Application Programming Interfaces

*** Graph-based Analytics**

- Incremental Analysis
- F2F Analysis
- Community Analysis
- Navigation
- Validation Overlay
- Decentralized Learning

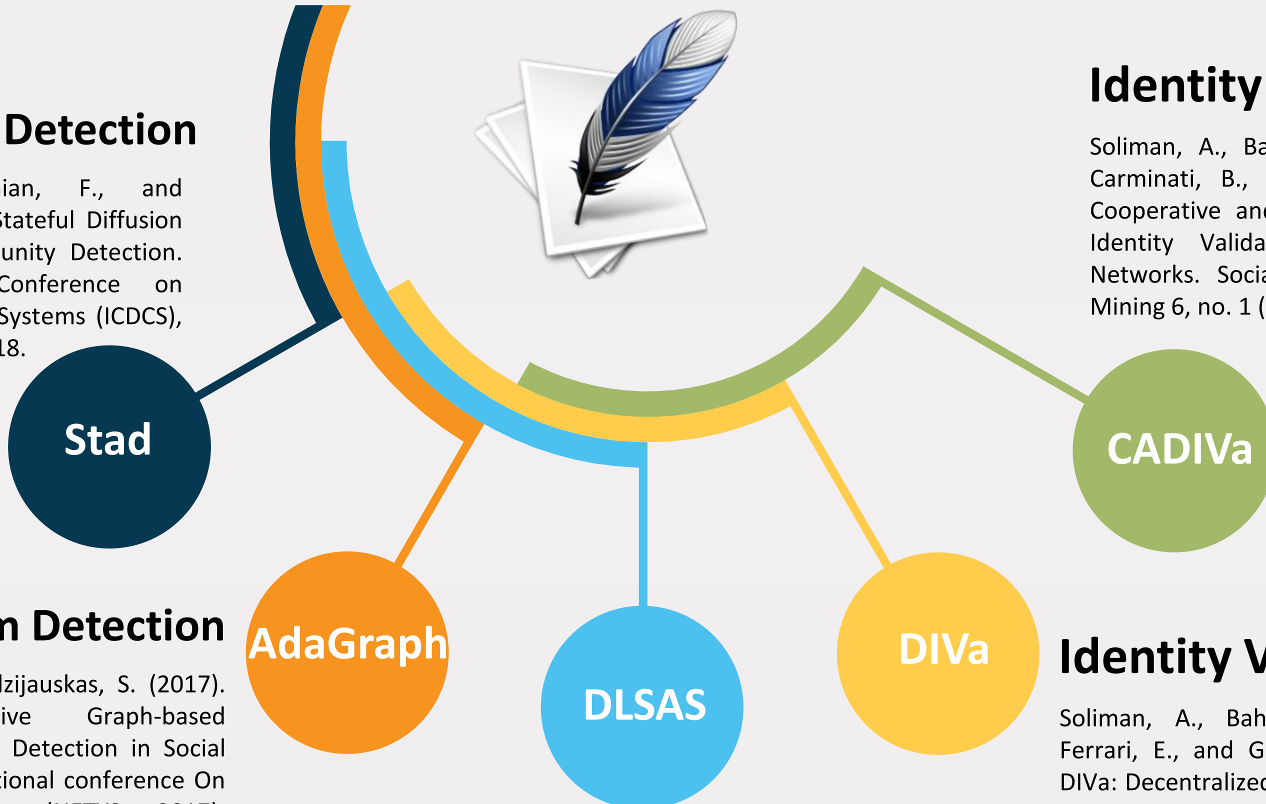
- Trust Management
- Pub/sub.
- Distributed Search
- Messaging
- …

Social Network Support

- Distributed Storage
- P2P Overlay Management
- Physical Communication Network

Distributed Network Support

Our Contributions



Community Detection

Soliman, A., Rahimian, F., and Girdzijauskas, S. Stad: Stateful Diffusion for Linear Time Community Detection. 38th International Conference on Distributed Computing Systems (ICDCS), pp. 1074-1085. IEEE, 2018.

Identity Validation

Soliman, A., Bahri, L., Girdzijauskas, S., Carminati, B., and Ferrari, E. CADIVa: Cooperative and Adaptive Decentralized Identity Validation Model for Social Networks. Social Network Analysis and Mining 6, no. 1 (2016): 1-22.

Spam Detection

Soliman, A. and Girdzijauskas, S. (2017). AdaGraph: Adaptive Graph-based Algorithms for Spam Detection in Social Networks. In International conference On Networked Systems (NETYS 2017), Springer, pages 338-354.

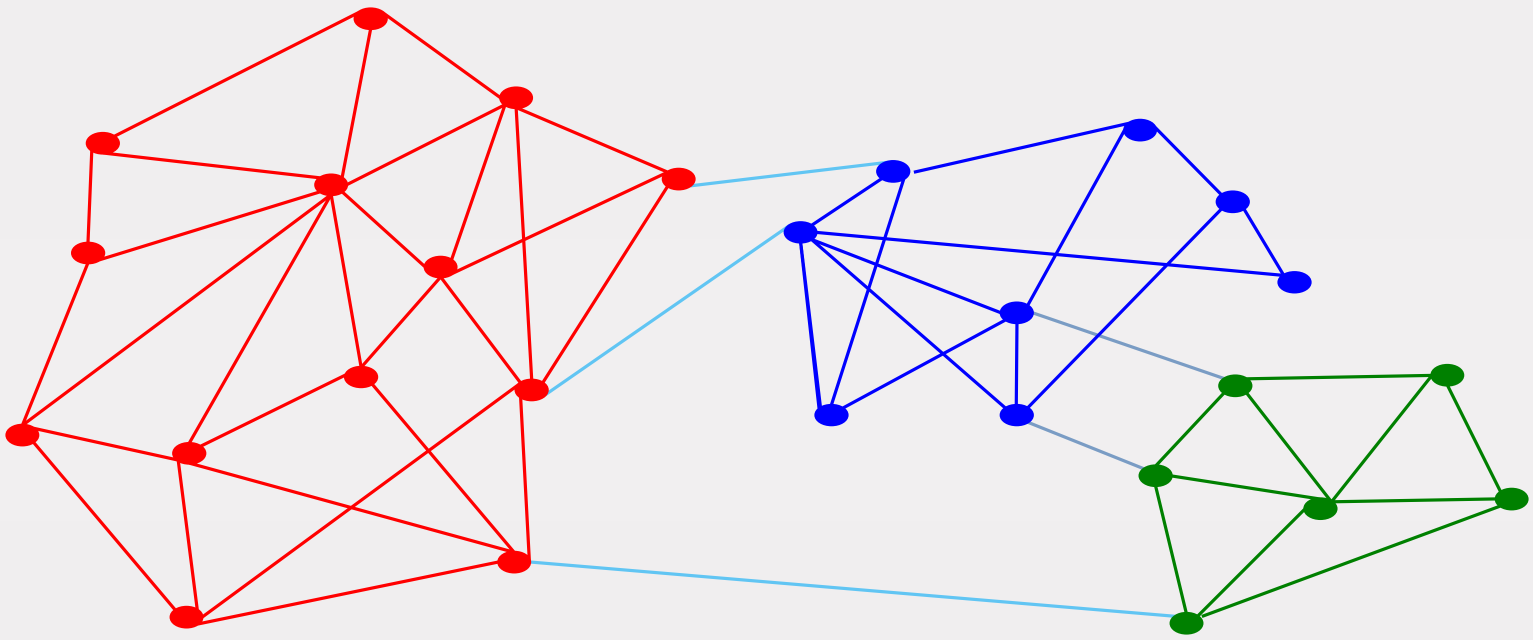
Validation Mechanism

Soliman, A. and Girdzijauskas, S. (2016). DLSAS: Distributed Large-Scale Anti-Spam Framework for Decentralized Online Social Networks. Invited paper in the 2nd IEEE International Conference on Collaboration and Internet Computing, IEEE, pages 363-372.

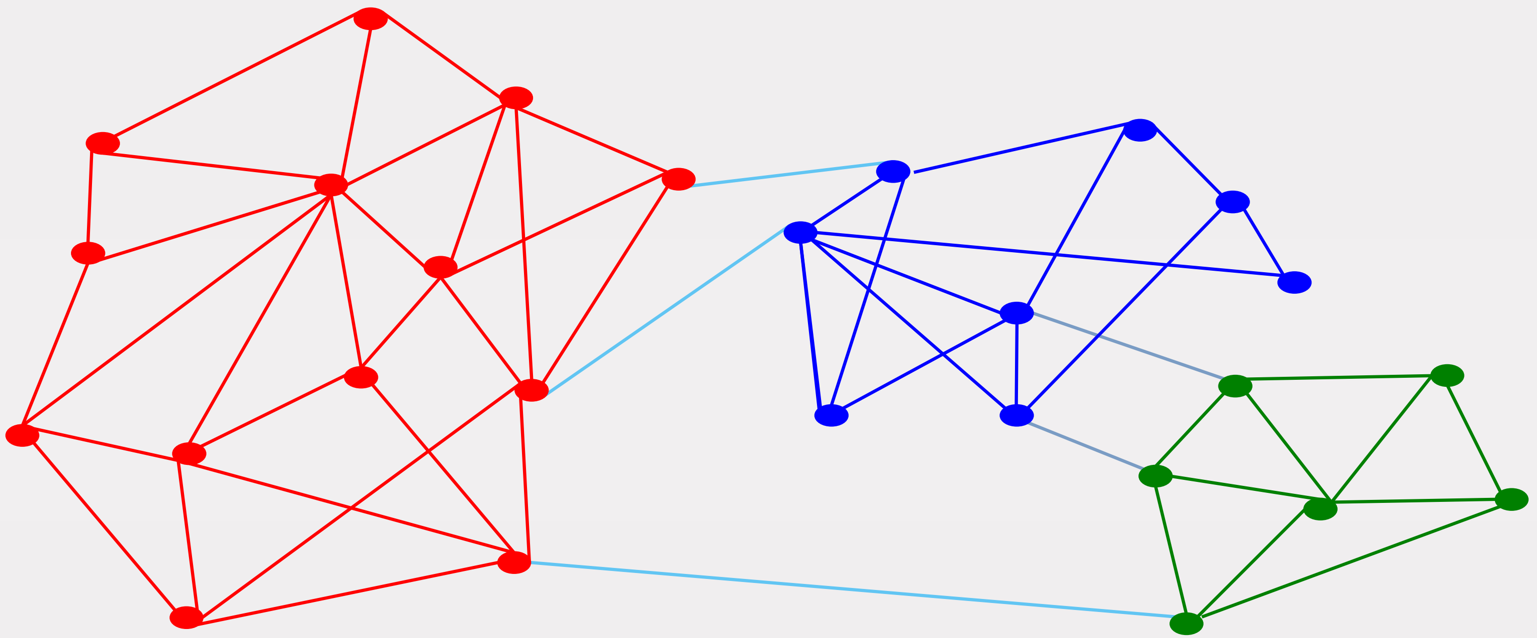
Identity Validation

Soliman, A., Bahri, L., Carminati, B., Ferrari, E., and Girdzijauskas, S. (2015). DIVa: Decentralized Identity Validation for Social Networks. In International Conference on Advances in Social Network Analysis and Mining (ASONAM), 2015 IEEE/ACM, pages 383-391.

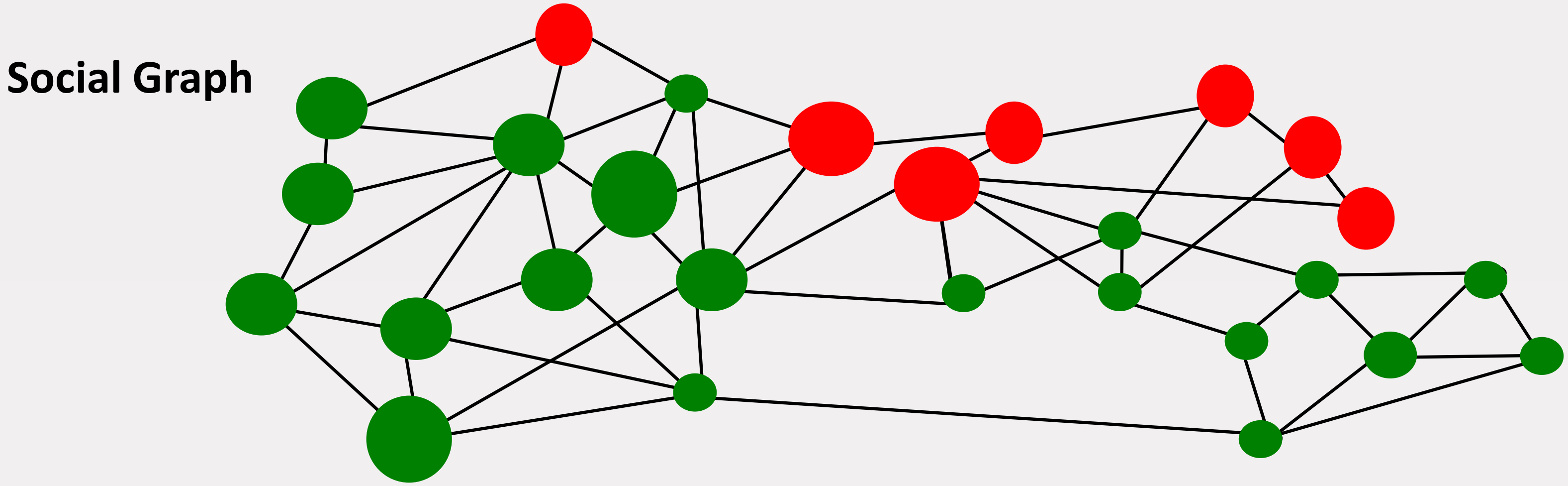
Stad: Performing Adaptive Diffusion for Community Detection



Stad: Performing Adaptive Diffusion for Community Detection

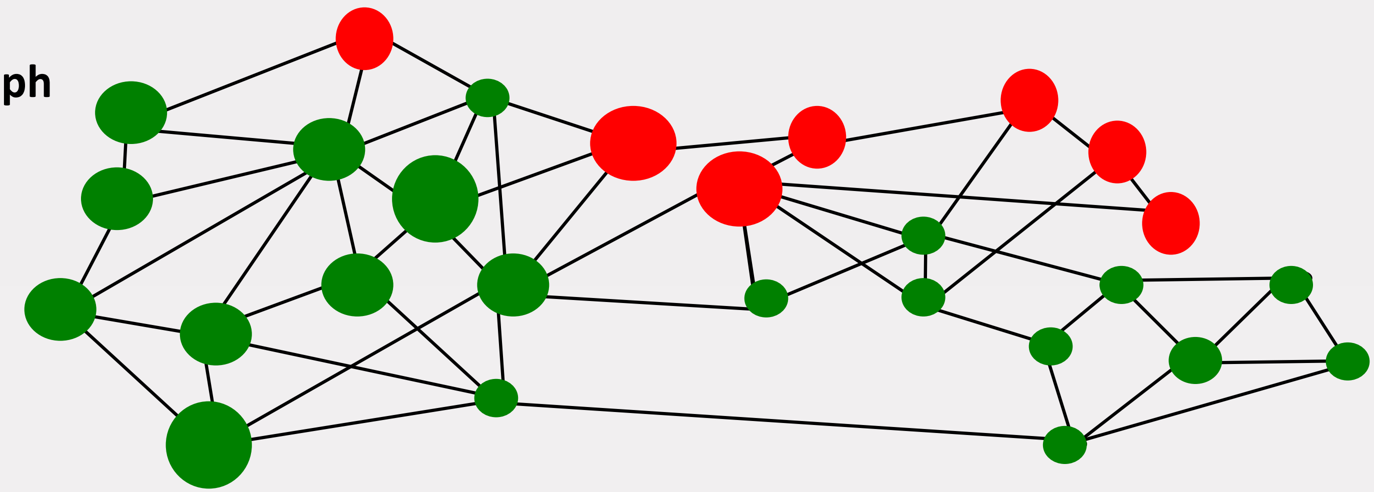


AdaGraph: Spam Detection using Graph Clustering



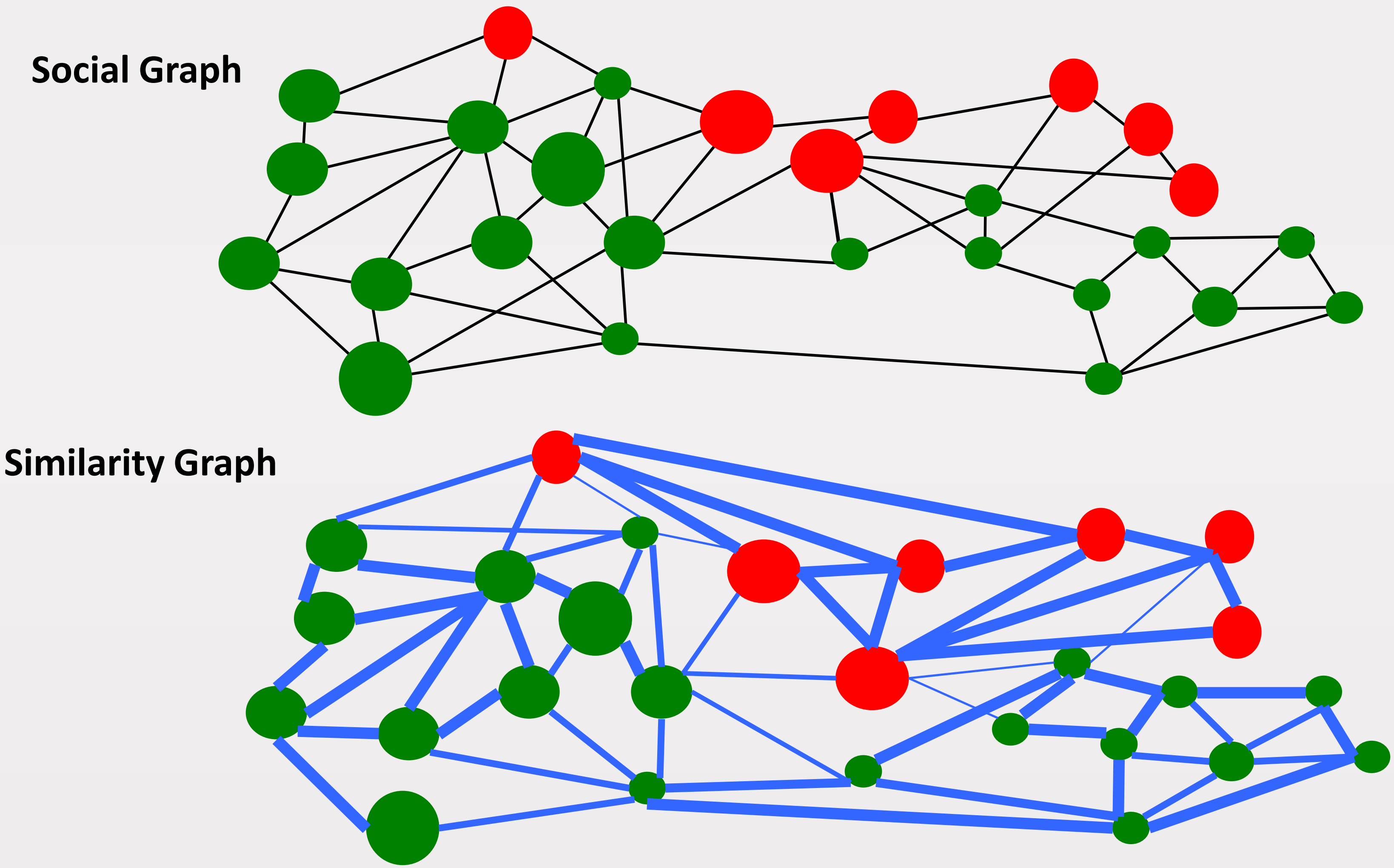
AdaGraph: Spam Detection using Graph Clustering

Social Graph

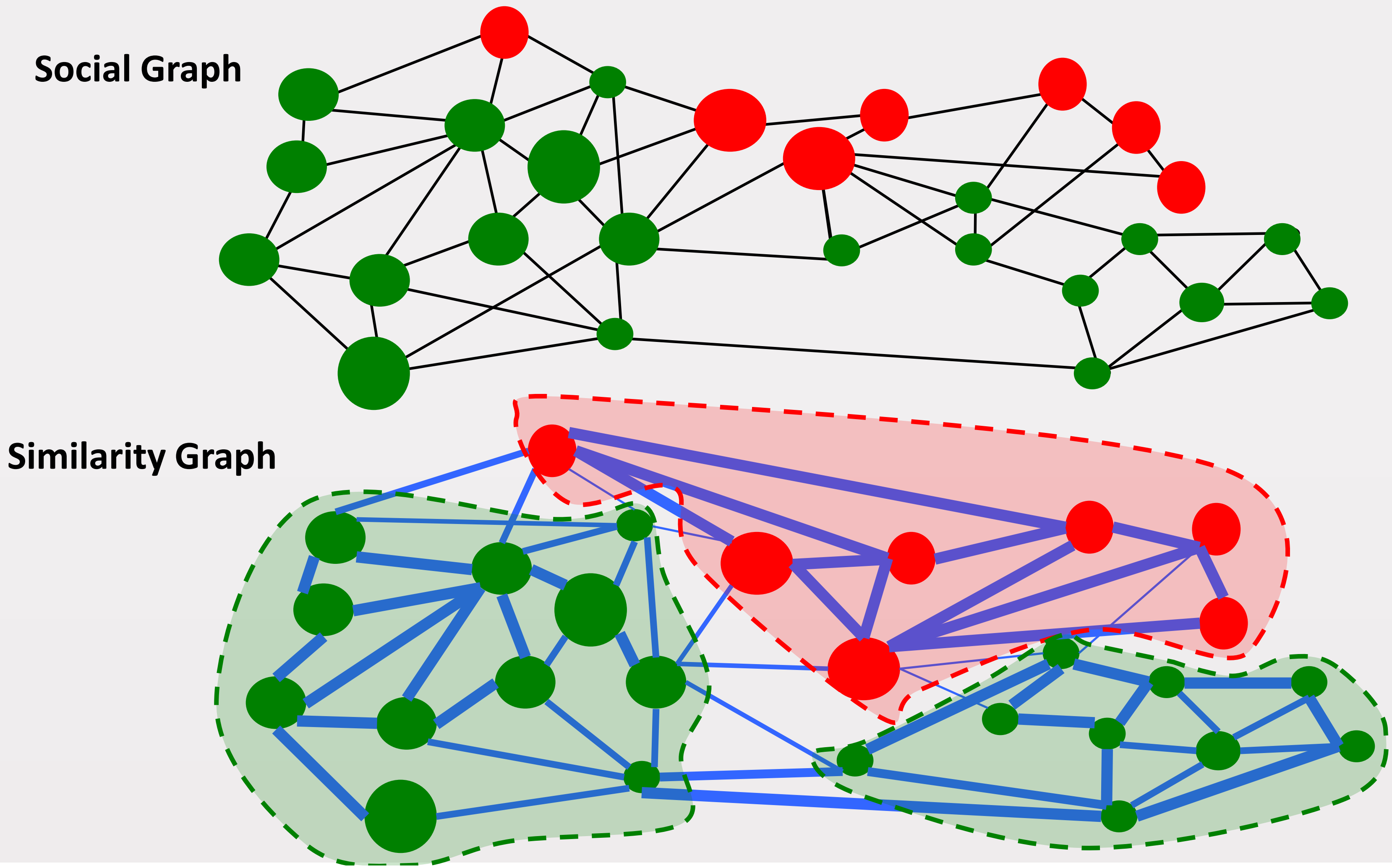


Similarity Graph

AdaGraph: Spam Detection using Graph Clustering

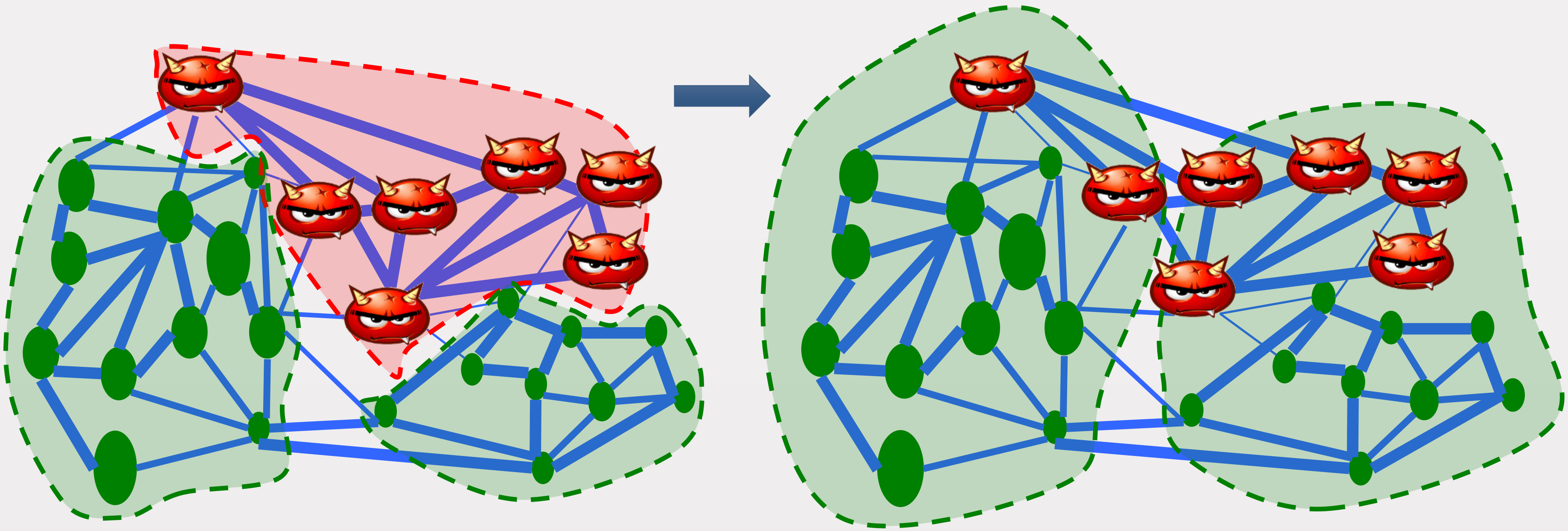


AdaGraph: Spam Detection using Graph Clustering



AdaGraph and Adversarial Manipulations

Malicious nodes collaborate with each other to falsify detected communities.

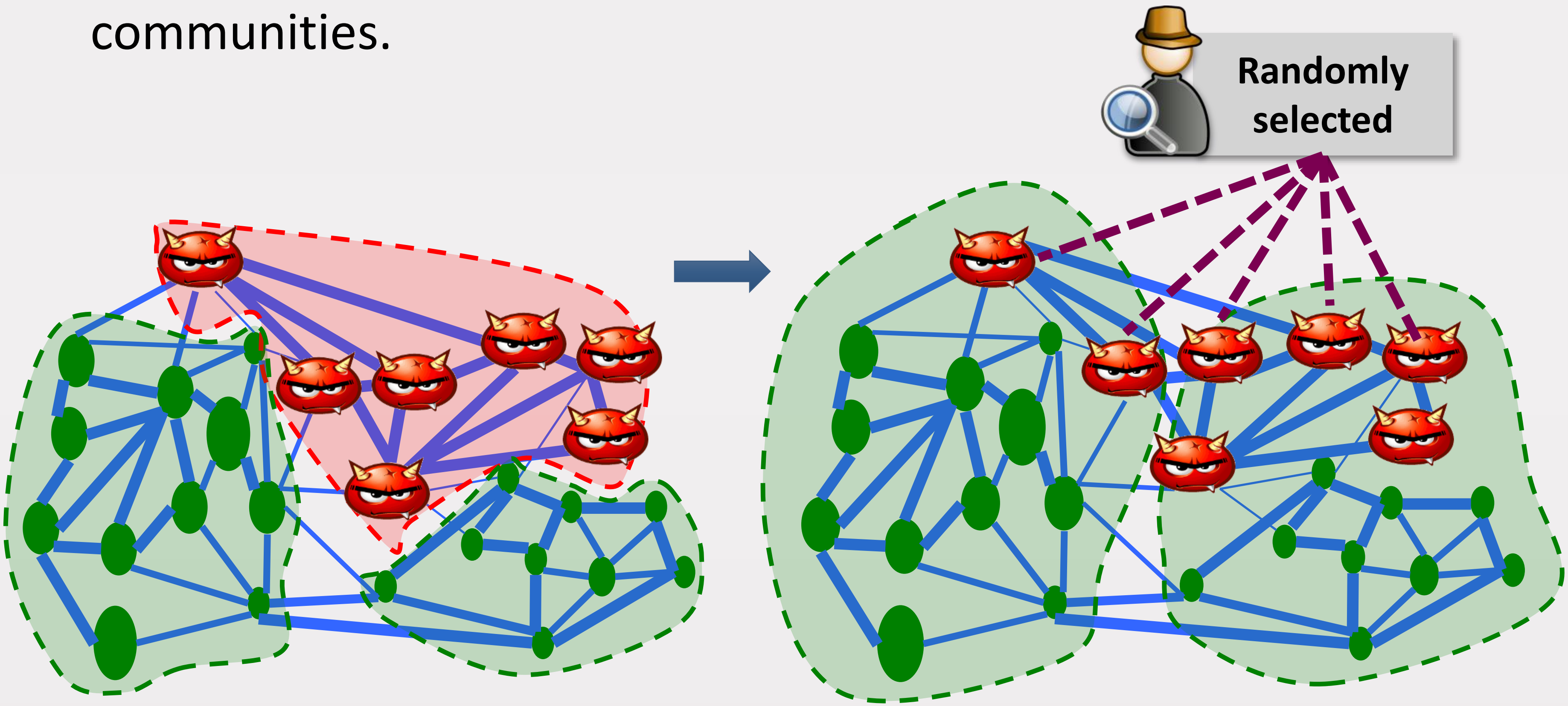


(a) Targeted detected communities

(b) Incorrect community memberships

AdaGraph and Adversarial Manipulations

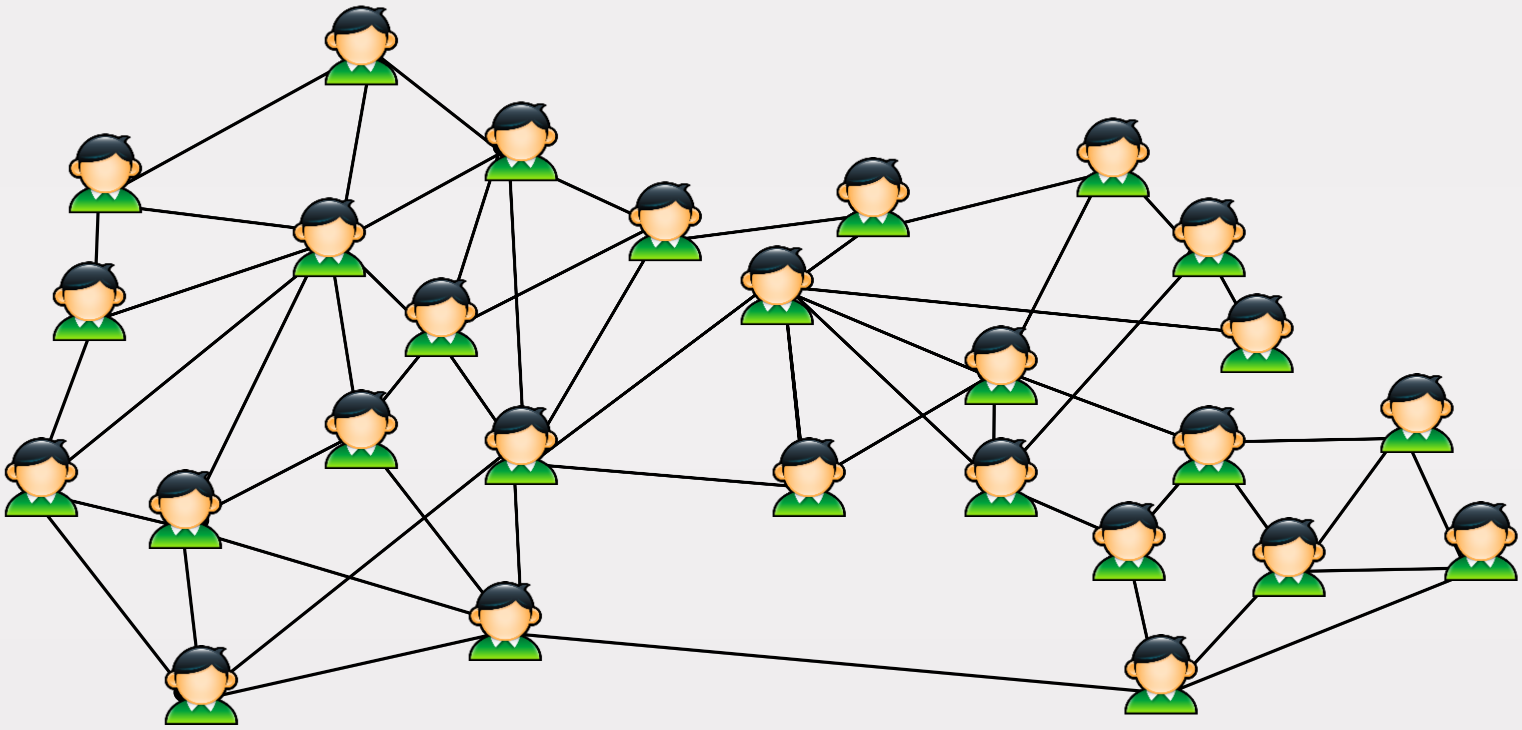
Malicious nodes collaborate with each other to falsify detected communities.



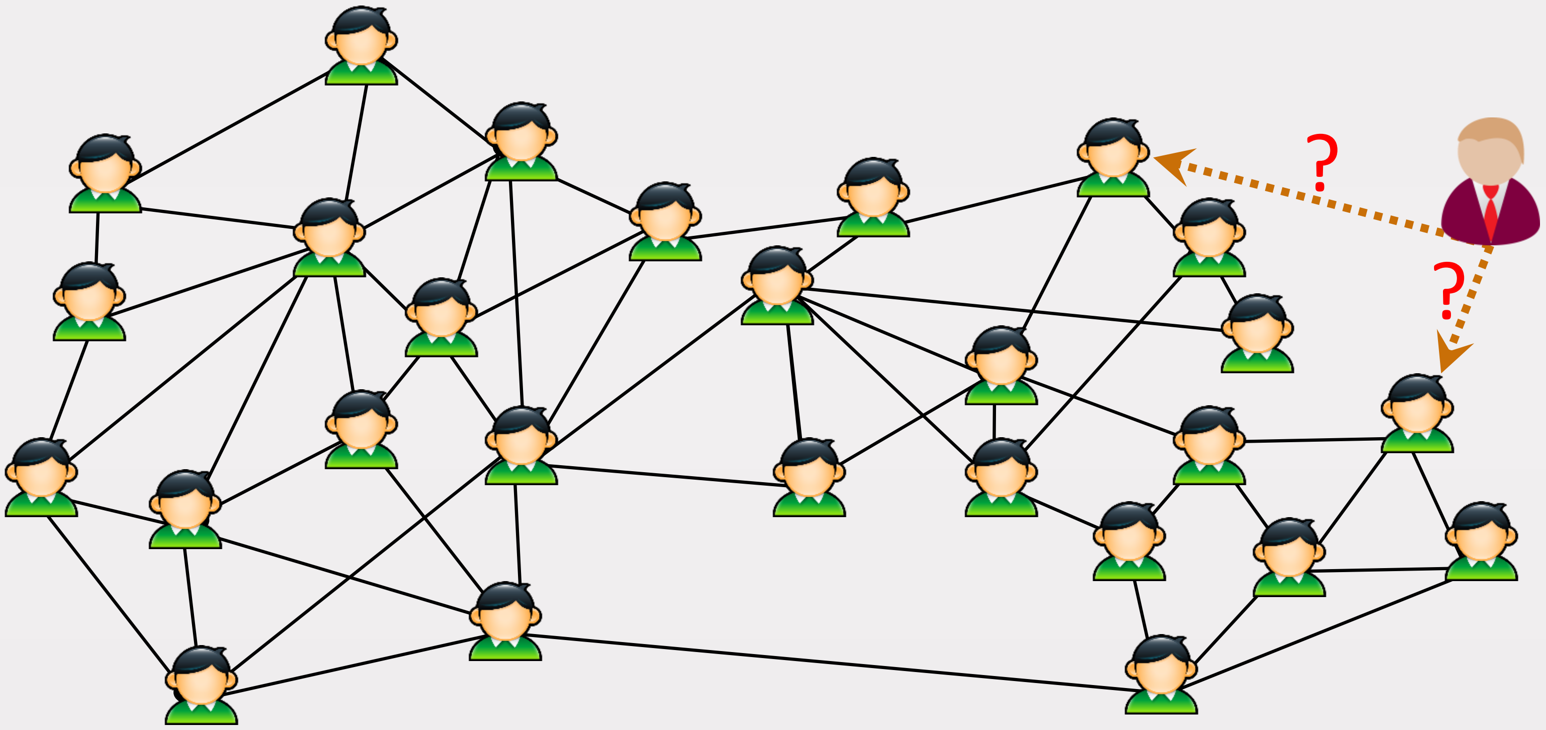
(a) Targeted detected communities

(b) Incorrect community memberships

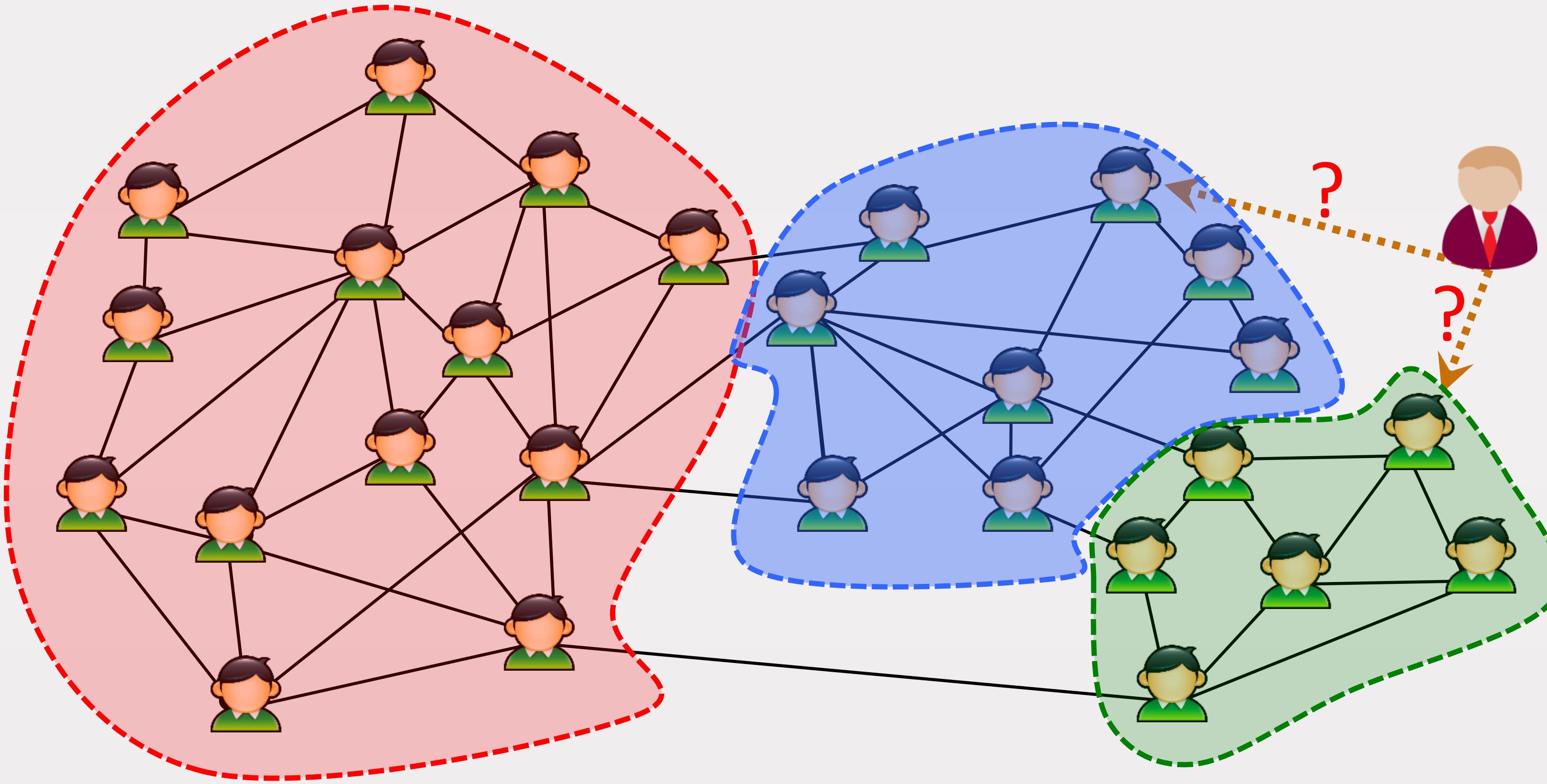
DIVa & CADIVa: How much trustworthy a profile is?



DIVa & CADIVa: How much trustworthy a profile is?



DIVa & CADIVa: How much trustworthy a profile is?



List of Publications

1. Soliman, A. , Rahimian, F., and Girdzijauskas, S. (2018). Stad: Stateful Diffusion for Linear Time Community Detection. 38th International Conference on Distributed Computing Systems (ICDCS), pp. 1074-1085. IEEE, 2018.
2. Soliman, A. and Girdzijauskas, S. (2017). AdaGraph: Adaptive Graph-based Algorithms for Spam Detection in Social Networks. In International conference On Networked Systems (NETYS 2017), Springer, pages 338-354.
3. Soliman, A. and Girdzijauskas, S. (2016). DLSAS: Distributed Large-Scale Anti-Spam Framework for Decentralized Online Social Networks. Invited paper in the 2nd IEEE International Conference on Collaboration and Internet Computing, IEEE, pages 363-372.
4. Soliman, A., Bahri, L., Girdzijauskas, S., Carminati, B., and Ferrari, E. CADIVa: Cooperative and Adaptive Decentralized Identity Validation Model for Social Networks. Social Network Analysis and Mining 6, no. 1 (2016): 1-22.
5. Bahri, L., Soliman, A., Squillaci, J., Carminati, B., and Ferrari, E. and Girdzijauskas, S. (2016). Beat the DIVa -Decentralized Identity Validation for Online Social Networks. Demo paper in the 32nd IEEE International Conference on Data Engineering.
6. Soliman, A., Bahri, L., Carminati, B., Ferrari, E., and Girdzijauskas, S. (2015). DIVa: Decentralized Identity Validation for Social Networks. In International Conference on Advances in Social Network Analysis and Mining (ASONAM), 2015 IEEE/ACM, pages 383-391.

Graph-based Analytics for Decentralized Online Social Networks

Thanks ☺