SybilFuse: Combining Local Attributes with Global Structure to Perform Robust Sybil Detection

Peng Gao[1], Binghui Wang[2], Neil Zhenqiang Gong[2], Sanjeev R. Kulkarni[1], Prateek Mittal[1]

[1]*Princeton University*
[2]*Iowa State University*

Motivation

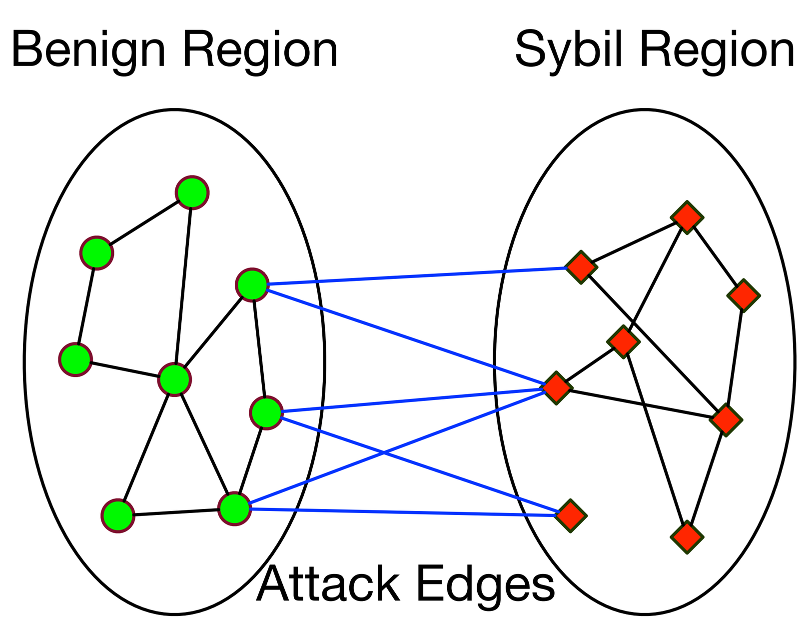


Fig.1: Sybil attack scenario

- Sybil attacks are poisonous and wide spread in online social networks such as Facebook and Twitter.
- The attacker can leverage Sybil accounts to:
  - Disrupt democratic election
  - Influence financial market
  - Propagate social malware
  - Disseminate scams
  - Carry out phishing attacks
  - Compromise user privacy

State-of-the-art

- Research proposes to mitigate Sybil attacks using *social network-based trust relationships*, i.e., structure-based approaches:
  - SybilGuard
  - SybilLimit
  - SybilInfer
  - SybilRank
  - Criminal Account Inference
  - SybilBelief
  - SybilSCAR
  - Integro
- Limitations:
  - Strong-trust assumptions are not necessarily satisfied.
  - Fail to leverage rich local information.

Contribution

- We propose SybilFuse, a framework that combines local attributes with global structure to perform Sybil detection.
- SybilFuse relaxes the assumptions in existing approaches.

Framework Overview

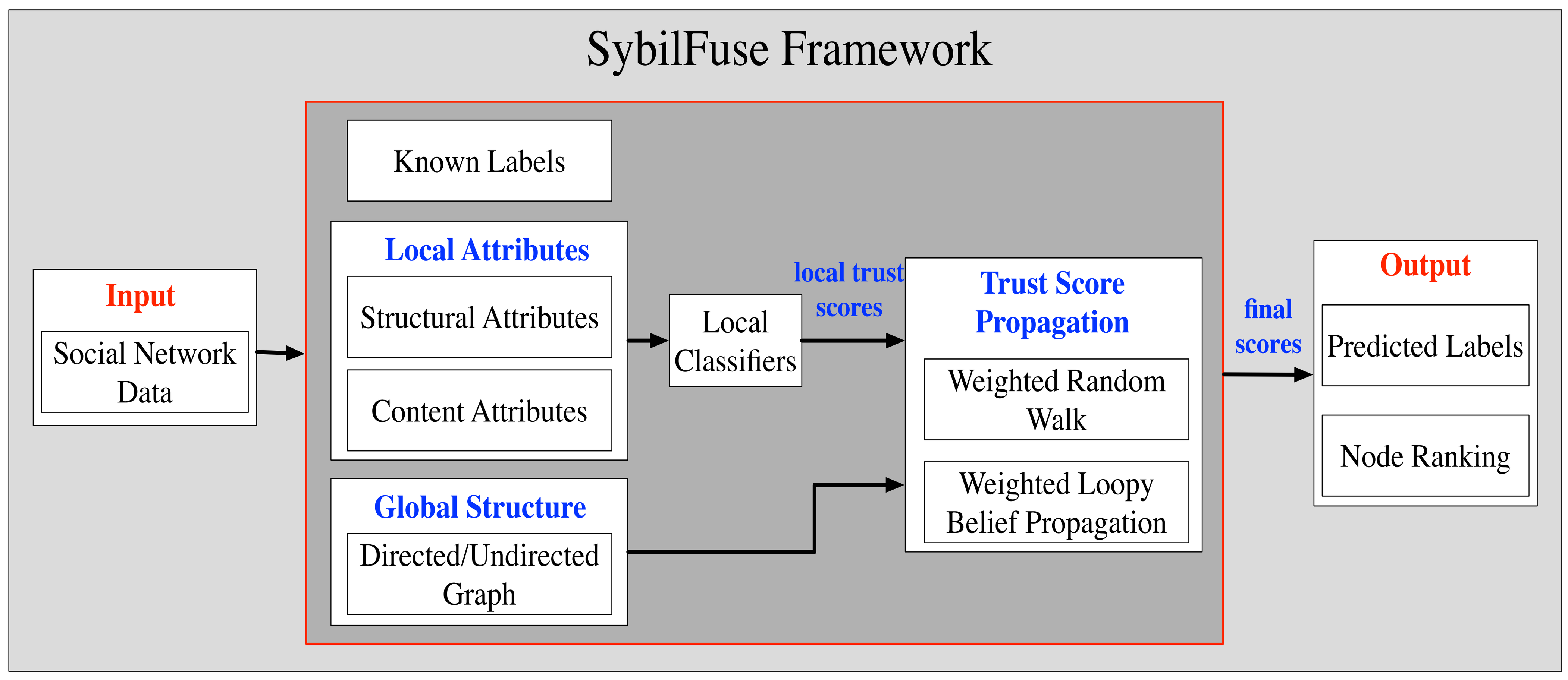


Fig.2: The SybilFuse framework

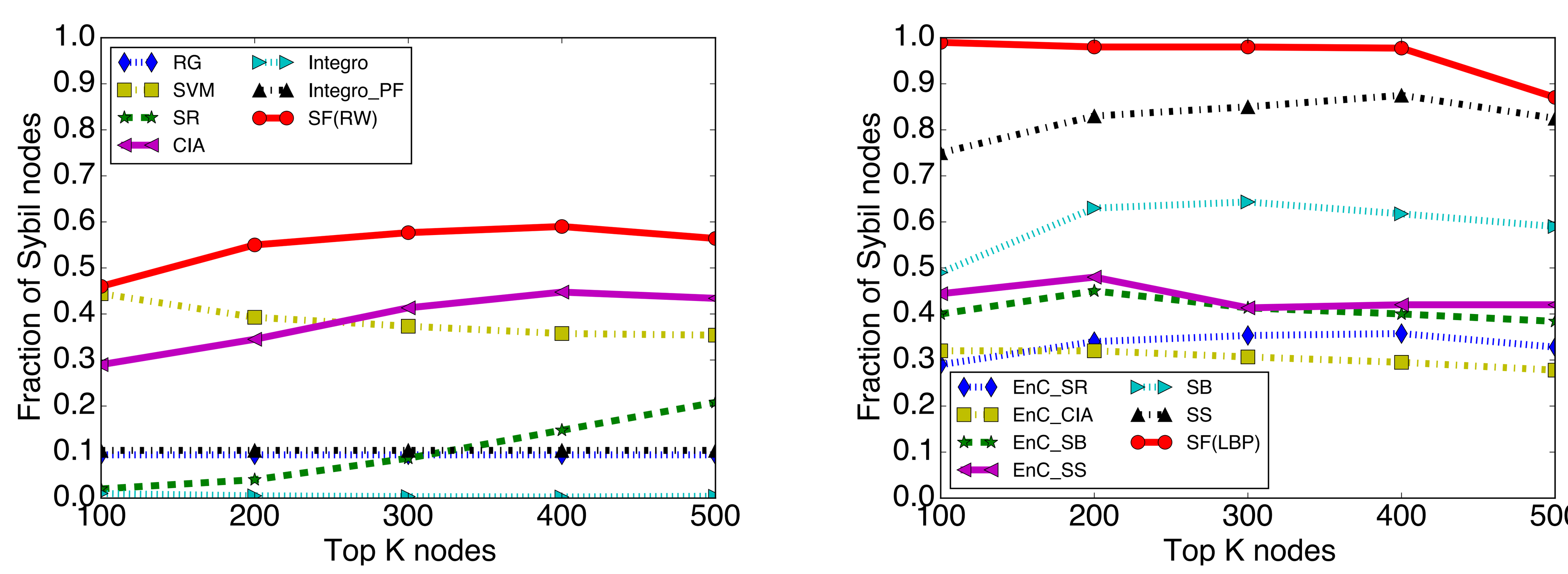
- Local trust score computation
  - Node trust score: quantifies the probability of a node to be a benign node $S_v \in [0, 1]$
  - Edge trust score: quantifies the probability of an edge to be a non-attack edge $S_{u,v} \in [0, 1]$
- Weighted trust score propagation
  - Weighted random walk

$$S^{(i)}(v) = \sum_{(u,v)\in E} S^{(i-1)}(u) \frac{S_{u,v}}{\sum_{(u,w)\in E} S_{u,w}}$$

  - Weighted loopy belief propagation

$$\psi_v(X_v) = \begin{cases} S_v & \text{if } X_v = 1 \\ 1 - S_v & \text{if } X_v = -1 \end{cases} \quad \psi_{u,v}(X_u, X_v) = \begin{cases} S_{u,v} & \text{if } X_u X_v = 1 \\ 1 - S_{u,v} & \text{if } X_u X_v = -1 \end{cases}$$

Labeled Twitter Evaluation



(a) Random walk-based approaches

(b) LBP-based approaches and ensemble methods

Fig.3: Fraction of Sybils among top K ranked nodes. The dataset contains 8167 nodes and 54146 edges, with verified 809 Sybil nodes. The number of attack edges is 40001, namely, 49 attack edges on average per Sybil)

- Training local SVM classifiers:
  - Incoming requests accepted ratio

$$Req_{in}(v) = \frac{|In(v) \cap Out(v)|}{|In(v)|}$$

  - Outgoing requests accepted ratio

$$Req_{out}(v) = \frac{|In(v) \cap Out(v)|}{|Out(v)|}$$

  - Local clustering coefficients

$$CC(v) = \frac{|\{(i,j) : i,j \in Nei(v), (i,j) \in E\}|}{|Nei(v)|\,(|Nei(v)| - 1)}$$

- Evaluation results:
  - SF(RW) achieves the best performance among all random walk-based approaches.
  - SF(LBP) achieves the best performance among all evaluated approaches (>98% Sybil ranking up to top 400 nodes).