# Inferring Degrees from Incomplete Networks and Nonlinear Dynamics

**C. Jiang**, J. Gao, M. Magdon-Ismail

{jiangc4,gaoj8}@rpi.edu, magdon@cs.rpi.edu

Department of Computer Science,
Rensselaer Polytechnic Institution, Troy, NY


# Problem

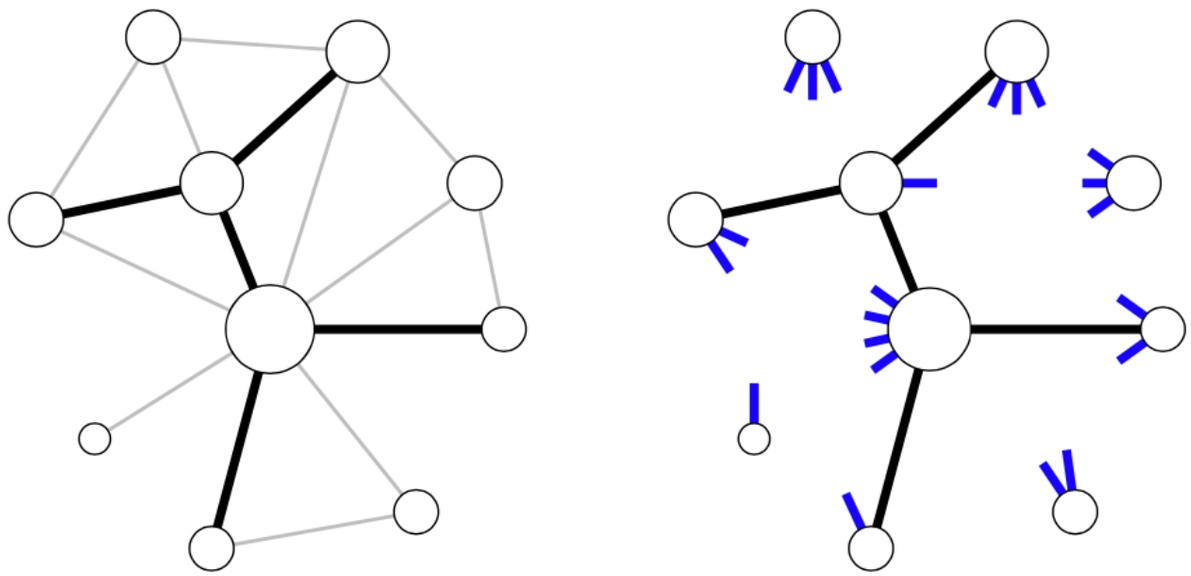
**Given:**

- Incomplete network
- Noisy equilibrium states
- Dynamical system

**Goals:**

- Infer missing degrees
- Link prediction
- Robustness (sampling, measurement error, model misspecification)

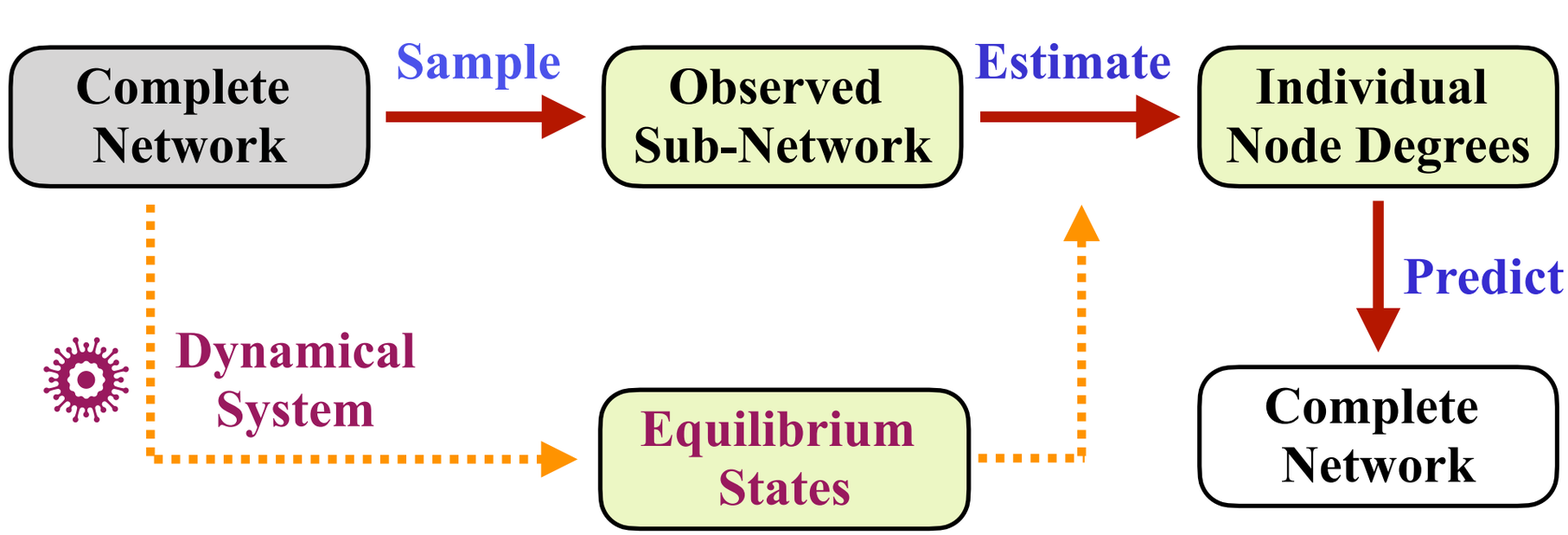
- **Ecology Networks**: abundances of plants
- **Regulatory Networks**: expression levels of genes
- **Epidemic Networks**: infection rates of person



# Prior Works

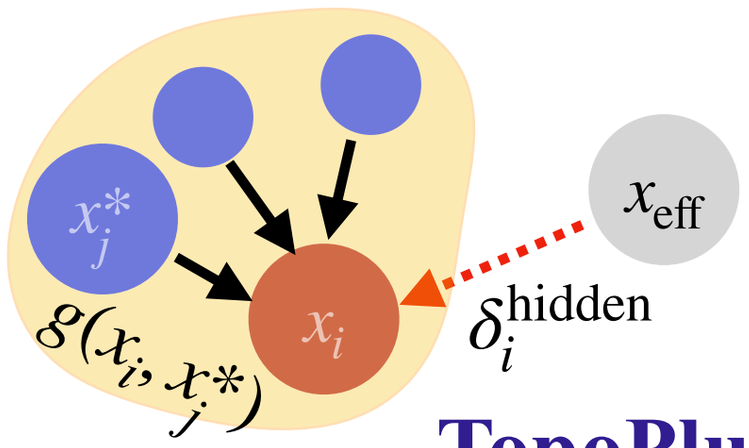
- **Degree distribution, average degree, network size** (BF1993; STS2008; KLS2011; KBM2012; GR2008; DKS2014; F1980; SW2005; AKM2009; RT2012; ZKS2015)

- **Individual node degrees** (GK2017)



# Main idea: connecting topology to equilibria

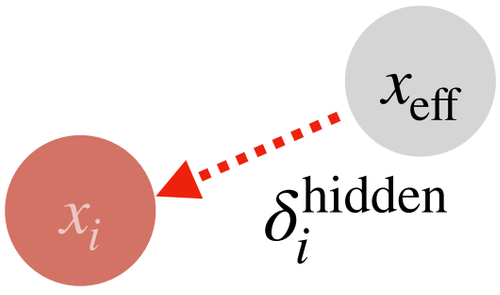
GBB2016 (Nature)
JGM2020 (AAAI)




**TopoPlus**

**Equilibrium Condition**

$$f(x_i) + \sum_{j \in V} A_{ij}^{(s)} g(x_i, x_j^*) + \delta_i^{\text{hidden}} g(x_i, x_{\text{eff}}) = 0$$




**ZeroTopo**

**Mean-field Equilibrium Condition**

$$f(x_i) + \delta_i^{\text{hidden}} g(x_i, x_{\text{eff}}) = 0$$

# Experimental Setting

| Dynamics | Network | $n$ | $m$ | $\langle \delta \rangle$ |
|---|---|---|---|---|
| ⓐ Ecological | Plant Pollinator | 97 | 972 | 20.04 |
| ⓑ Regulatory | Genes Network | 662 | 1062 | 3.21 |
| ⓒ Epidemic | Facebook | 4039 | 88234 | 43.69 |

ⓐ $\dot{x}_i = B + x_i(1 - \frac{x_i}{K})(\frac{x_i}{C} - 1) + \sum_j \frac{x_i x_j A_{ij}}{D + E x_i + H x_j}$

ⓑ $\dot{x}_i = -B x_i^f + \sum_j A_{ij} R \frac{x_j^h}{x_j^h + 1}$

ⓒ $\dot{x}_i = -B x_i + \sum_j A_{ij} R (1 - x_i) x_j$

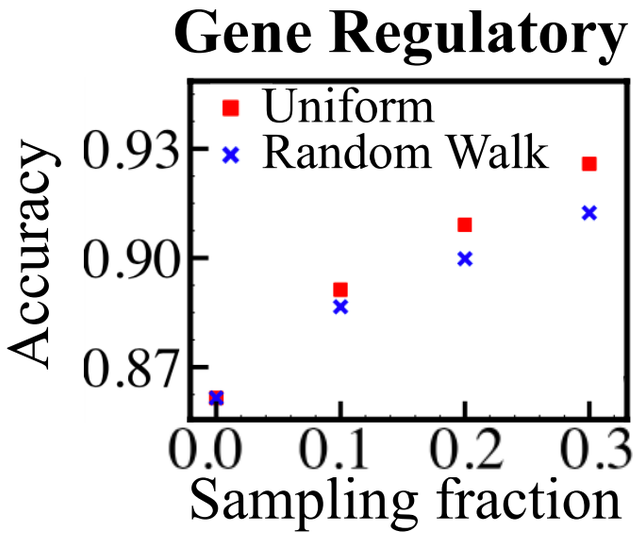
# Inferring Individual Node Degrees

| Network | Method (sampling fraction) | | |
|---|---|---|---|
| | ZeroTopo (**0%**) | TopoPlus (10%) | TopoPlus+***Round*** (10%) |
| Plant Pollinator | 57.7 | 62.0 | 62.7 |
| Genes | 86.2 | 87.1 | 87.8 |
| Facebook | 59.5 | 64.1 | 65.4 |

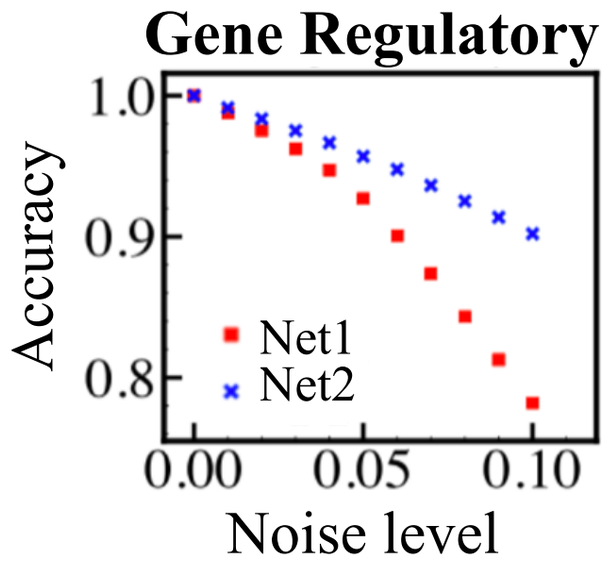
**Accuracy**: fraction of nodes with error $|\log(\hat{\delta}_i/\delta_i)| < 5\,\%$

# Robustness

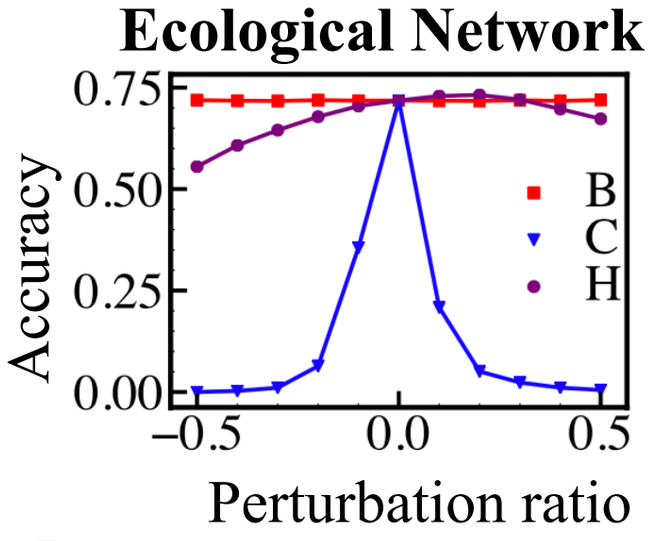
- **Sampling Method**




- **Measurement Error**




- **Model Error**



# Comparison: degree estimation

**On average 157% improvement over [GK2017]**

| **Network** / **Method** | Plant Pollinator | Genes | Facebook |
|---|---|---|---|
| [GK2017] | 5.10 | 19.96 | 9.00 |
| ZeroTopo (*w/* noise) | **12.63 (10%)** | **84.27(10%)** | **9.19 (9%)** |

# Degree estimation for link prediction

- **Our estimates: 41% improvement over observed degrees**
- **True degrees: 44% improvement over observed degrees**

| Method | AUC | | |
|---|---|---|---|
| | Plant Pollinator | Genes | Facebook |
| [BA1999] + Observed degrees | 59.90 | 58.00 | 69.44 |
| **[BA1999] + Our estimates** | **87.11** | **91.28** | **83.78** |
| [BA1999] + True degrees | 87.42 | 92.45 | 83.83 |

- All methods sample 1% of edges
- [BA1999] uses $s_{u,v} = \delta_u \delta_v$ for link prediction
- Our degree estimates with **TopoPlus** assumes no measurement error

# Conclusions

- Our degree estimates use equilibrium states
- Accurate estimates even without topology, unlike existing methods
- Robust to sampling method, measurement errors, and model errors
- Our degree estimates significantly enhance link prediction

# References


[AKM2009] Dimitris Achlioptas, Clauset, David Kempe, and Cristopher Moore. On the bias of traceroute sampling: or, power-law degree distri- butions in regular graphs. *Journal of the ACM (JACM)*, 56(4):21, 2009.

[BA1999] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[BF1993] John Bunge and Michael Fitzpatrick. Estimating the number of species: a review. *Journal of the American Statistical Association*, 88(421):364–373, 1993.

[DKS2014] Anirban Dasgupta, Ravi Kumar, and Tamas Sarlos. On estimating the average degree. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 795–806. ACM, 2014.

[F1980] Ove Frank. Estimation of the number of vertices of different degrees in a graph. *Journal of Statistical Planning and Inference*, 4(1):45–50, 1980.

[GK2017] Apratim Ganguly and Eric D Kolaczyk. Estimation of vertex degrees in a sampled network. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pages 967–974. IEEE, 2017.

[GBB2016] Jianxi Gao, Baruch Barzel, and Albert-Laszlo Barabasi. Universal resilience patterns in complex networks. *Nature*, 530(7590):307, 2016.

[GR2008] Oded Goldreich and Dana Ron. Approximating average parameters of graphs. *Random Structures & Algorithms*, 32(4):473–493, 2008.

[JGM2020] Chunheng Jiang, Jianxi Gao, and Malik Magdon-Ismail. True nonlinear dynamics from incomplete networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 131–138, 2020.

# References


[KLS2011] Liran Katzir, Edo Liberty, and Oren Somekh. Estimating sizes of social networks via biased sampling. In *Proceedings of the 20th International Conference on World Wide Web*, pages 597–606. ACM, 2011.

[KBM2012] Maciej Kurant, Carter T Butts, and Athina Markopoulou. Graph size estimation. *arXiv preprint arXiv:1210.0460*, 2012.

[RT2012] Bruno Ribeiro and Don Towsley. On the estimation accuracy of degree distributions from graph sampling. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 5240–5247. IEEE, 2012.

[STS2008] Michael PH Stumpf, Thomas Thorne, Eric de Silva, Ronald Stewart, Hyeong Jun An, Michael Lappe, and Carsten Wiuf. Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences*, 105(19):6959–6964, 2008.

[SW2005] Michael PH Stumpf and Carsten Wiuf. Sampling properties of random graphs: the degree distribution. *Physical Review E*, 72(3):036118, 2005.

[ZKS2015] Yaonan Zhang, Eric D Kolaczyk, Bruce D Spencer, et al. Estimating network degree distributions under sampling: An inverse problem, with applications to monitoring social media networks. *The Annals of Applied Statistics*, 9(1):166–199, 2015.