A Statistician Worries About Random Network Modeling

Norm Matloff
Dept. of Computer Science
University of California, Davis

May 18, 2009

Where I'm Coming From

Question at hand:

What is the state of the art on networks research in your discipline?

Where I'm Coming From

Question at hand:

> *What is the state of the art on networks research in your discipline?*

My computer science affiliation might be misleading.

Where I'm Coming From

Question at hand:

What is the state of the art on networks research in your discipline?

My computer science affiliation might be misleading.

- Dissertation in pure math (statistical physics, but ancient history, 1975).

Where I'm Coming From

Question at hand:

What is the state of the art on networks research in your discipline?

My computer science affiliation might be misleading.

- Dissertation in pure math (statistical physics, but ancient history, 1975).
- Early career was as a statistics professor.

Where I'm Coming From

Question at hand:

What is the state of the art on networks research in your discipline?

My computer science affiliation might be misleading.

- Dissertation in pure math (statistical physics, but ancient history, 1975).
- Early career was as a statistics professor.
- UCD CS Dept. since 1983, but much of my work is statistical.

Where I'm Coming From

Question at hand:

> *What is the state of the art on networks research in your discipline?*

My computer science affiliation might be misleading.

- Dissertation in pure math (statistical physics, but ancient history, 1975).
- Early career was as a statistics professor.
- UCD CS Dept. since 1983, but much of my work is statistical.
- I don't really consider random networks to be CS, though there are CS applications.

Where I'm Coming From

Question at hand:

What is the state of the art on networks research in your discipline?

My computer science affiliation might be misleading.

- Dissertation in pure math (statistical physics, but ancient history, 1975).
- Early career was as a statistics professor.
- UCD CS Dept. since 1983, but much of my work is statistical.
- I don't really consider random networks to be CS, though there are CS applications.

So, in this context I take "my discipline" to be statistics, not CS.

Where I'm Coming From

Question at hand:

> *What is the state of the art on networks research in your discipline?*

My computer science affiliation might be misleading.

- Dissertation in pure math (statistical physics, but ancient history, 1975).
- Early career was as a statistics professor.
- UCD CS Dept. since 1983, but much of my work is statistical.
- I don't really consider random networks to be CS, though there are CS applications.

So, in this context I take "my discipline" to be statistics, not CS.

And as a statistician, many aspects of the current state of random network models worry me.

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

- Prediction?

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

- Prediction?
  - Next new alliance [works by Maoz etc.].

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

- Prediction?
  - Next new alliance [works by Maoz etc.].
  - Adolescents at risk of running away [Borgatti *et al*, 2009].

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

- Prediction?
  - Next new alliance [works by Maoz etc.].
  - Adolescents at risk of running away [Borgatti *et al*, 2009].
- Understanding?

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

- Prediction?
  - Next new alliance [works by Maoz etc.].
  - Adolescents at risk of running away [Borgatti *et al*, 2009].
- Understanding?
  - Phase change may have useful “tipping point” interpretations.

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

- Prediction?
  - Next new alliance [works by Maoz etc.].
  - Adolescents at risk of running away [Borgatti *et al*, 2009].
- Understanding?
  - Phase change may have useful “tipping point” interpretations.
  - Effects of covariates on link formation.

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

- Prediction?
  - Next new alliance [works by Maoz etc.].
  - Adolescents at risk of running away [Borgatti *et al*, 2009].
- Understanding?
  - Phase change may have useful “tipping point” interpretations.
  - Effects of covariates on link formation.
  - M/G/1 queue dep. on variance of service time.

What Are the Goals?

Hard to choose research directions unless we know what the goals are.

- Prediction?
  - Next new alliance [works by Maoz etc.].
  - Adolescents at risk of running away [Borgatti *et al*, 2009].
- Understanding?
  - Phase change may have useful “tipping point” interpretations.
  - Effects of covariates on link formation.
  - M/G/1 queue dep. on variance of service time.

Let’s look at some issues with these points in mind.

Model Robustness Issues

Contrast classical statistical modeling, say regression analysis, with that of random networks.

Model Robustness Issues

Contrast classical statistical modeling, say regression analysis, with that of random networks.

Regression:

Model Robustness Issues

Contrast classical statistical modeling, say regression analysis, with that of random networks.

Regression:

- Good robustness for Prediction (due to minimizing prediction some of squares).

Model Robustness Issues

Contrast classical statistical modeling, say regression analysis, with that of random networks.

Regression:

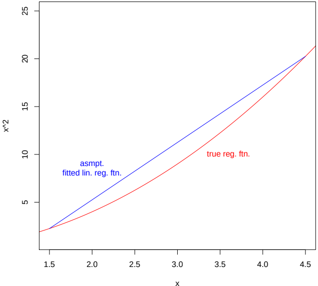
- Good robustness for Prediction (due to minimizing prediction some of squares).
- Fairly good robustness for Understanding. If true regression function is monotonic with only one predictor variable, β_i will be "about right."

Model Robustness Issues

Contrast classical statistical modeling, say regression analysis, with that of random networks.

Regression:

- Good robustness for Prediction (due to minimizing prediction some of squares).
- Fairly good robustness for Understanding. If true regression function is monotonic with only one predictor variable, β_i will be "about right."



Model Robustness Issues (cont'd.)

Random graphs:

Robustness not much explored yet.

Model Robustness Issues (cont'd.)

Random graphs:

Robustness not much explored yet.

- Common situation [Aielo *et al*, 2002]:

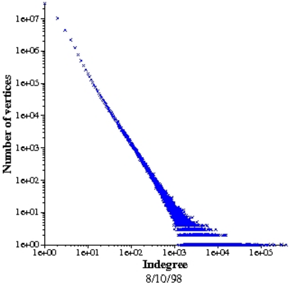


Figure 2: *The number of vertices for each possible indegree for the call graph of a typical day.*

Model Robustness Issues (cont'd.)

Random graphs:

Robustness not much explored yet.

- Common situation [Aielo *et al*, 2002]:

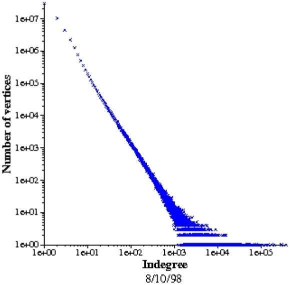


Figure 2: *The number of vertices for each possible indegree for the call graph of a typical day.*

- Model problems in tail raises concern about both Prediction and Understanding.

Model Robustness Issues (cont'd.)

Random graphs:

Robustness not much explored yet.

- Common situation [Aielo *et al*, 2002]:

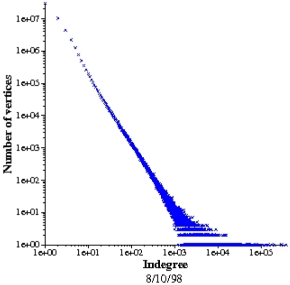


Figure 2: *The number of vertices for each possible indegree for the call graph of a typical day.*

- Model problems in tail raises concern about both Prediction and Understanding.
- Seemingly small change in assumptions can greatly change qualitative behavior [D'Souza *et al*, 2009].

Model Goodness of Fit (GOF) Assessment

Again, let's contrast regression and random network models.
Regression:

Model Goodness of Fit (GOF) Assessment

Again, let's contrast regression and random network models.
Regression:

- If desire formal testing, have a natural hierarchy of models (multivariate polynomial) for assessing GOF.

Model Goodness of Fit (GOF) Assessment

Again, let's contrast regression and random network models.
Regression:

- If desire formal testing, have a natural hierarchy of models (multivariate polynomial) for assessing GOF.
- Informal graphical assessment (residuals, nonparametric estimation) is simple and easily interpretable.

Model Goodness of Fit (GOF) Assessment

Again, let's contrast regression and random network models.
Regression:

- If desire formal testing, have a natural hierarchy of models (multivariate polynomial) for assessing GOF.
- Informal graphical assessment (residuals, nonparametric estimation) is simple and easily interpretable.
- Cross-validation (splitting data into training, assessment sets) is easy.

GOF Assessment (cont'd.)

Random networks:

GOF Assessment (cont'd.)

Random networks:

- Other than ERGM, formal complete tests lacking, maybe impossible.

GOF Assessment (cont'd.)

Random networks:

- Other than ERGM, formal complete tests lacking, maybe impossible.
- N often huge, so tests are meaningless anyway.

GOF Assessment (cont'd.)

Random networks:

- Other than ERGM, formal complete tests lacking, maybe impossible.
- N often huge, so tests are meaningless anyway.
- Graph subsets often behave differently from the full set [Stumpf *et al*, 2005], so cross-validation doesn't work.

Other Major Issues

Other Major Issues

- Hypothesis testing, confidence intervals:

Other Major Issues

- Hypothesis testing, confidence intervals:
  - In general, nonindependence of observations issue not solved, difficult to attack.

Other Major Issues

- Hypothesis testing, confidence intervals:
  - In general, nonindependence of observations issue not solved, difficult to attack.
  - The bootstrap, “the statistician’s Swiss army knife,” inoperable (subgraph, nonindependence problems).

Other Major Issues

- Hypothesis testing, confidence intervals:
  - In general, nonindependence of observations issue not solved, difficult to attack.
  - The bootstrap, “the statistician’s Swiss army knife,” inoperable (subgraph, nonindependence problems).
- Bias issues [Achlioptas, 2005] more common here than in classical statistics.

(One) Statistician's Conclusion

(One) Statistician's Conclusion

- Random networks could revitalize the moribund statistics profession.

(One) Statistician's Conclusion

- Random networks could revitalize the moribund statistics profession.
- Model fragility, GOF assessment issues are troubling.

(One) Statistician's Conclusion

- Random networks could revitalize the moribund statistics profession.
- Model fragility, GOF assessment issues are troubling.
- Potential for misunderstanding, and thus misuse, is signficant.