Graphical Models, Social Network Analysis, and Structural Modeling (in International Relations)

Langche Zeng
The George Washington University

GWIPP Public Policy Methods Workshop
March 29, 2004

Introduction/Overview

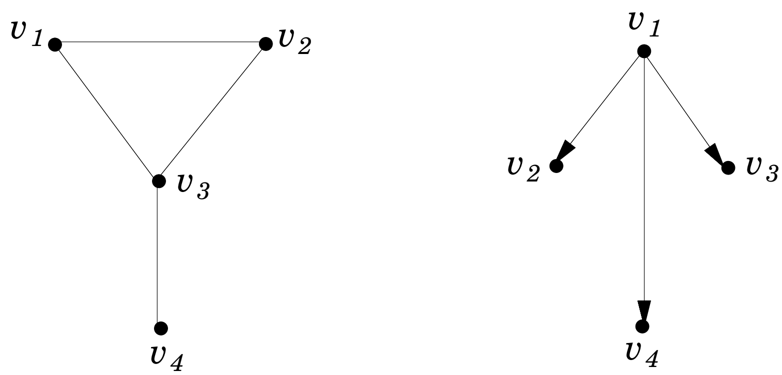
- Prediction and causal inference central pursuits of empirical work
- Understanding structural relationships critical to reliable prediction and causal inference. e.g. International events take place not in isolation, but within a network of intricate interdependence; structural properties of the system likely hold explanatory and predictive power
- Graphical methods and models (based on or aided with graph theory, a branch of mathematics) excellent tools for conceptualizing and representing relations and structures. e.g.:

  —relationships among social actors (individuals, states, organizations, etc.) e.g. the relation of "being in conflict with.".

  —relationships among variables in a model (causal structure; functional mapping)
- Different types of graphical models.. Important recent development:
  1. random graphs for relational data.
  2. causal graph theory.
- Talk focuses on one particular type of graphical model: Social Networks, with application to international conflict data. in particular:

  —graph theoretic measures of structural characteristics

  —modeling endogenous dependence structure using random graph model



Graphs

- A *graph* G is a pair $G = (V, E)$

  V: finite set of *vertices* or *nodes*. E: set of *edges*, (2-element subsets of V), representing a *relation* on V.

- symmetric relation $\rightarrow$ undirected edge (e.g. conflict involvement); asymmetric: directed edge (e.g. conflict initiation). if all edges of G directed, G is a digraph.

- matrix representation: X, such that the ijth element of X is 1 if there is a link from node i to j, 0 otherwise. For binary relation, X called the *adjacency matrix* of G. (uv is an edge of $G \rightarrow u$ and v are *adjacent*.)

- A graph is *complete* if all pairs of nodes are adjacent.

- subgraphs; cliques (complete subgraphs)

Graphical models

Combinations of particular types of "nodes" and "relations" give rise to various specific graphical models

- When nodes are variables in a system under study, edges are causal relations $\rightarrow$ causal graphs. (Pearl 2000). identification of causal effects; discovery of causal graph from data, etc.
- When nodes are variables, edges input–output relations $\rightarrow$ functional mapping, such as neural networks.
- When nodes are actors (states, dyads, individuals, organizations, etc.), and edges are relations among the actors $\rightarrow$ social networks

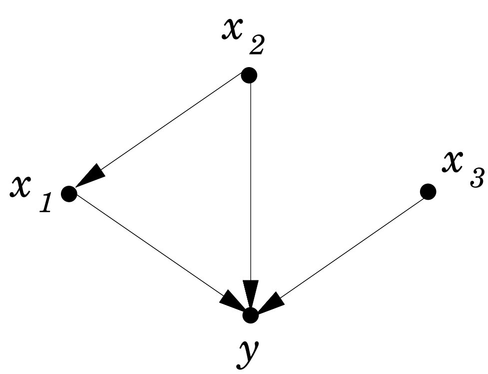
Causal graph and causal structure: Is the same model good for both prediction and causal inference?

- 99% of empirical work in political science uses the same (parametric) model for prediction and causal inference, and often for studying causal effects of *each* independent variable in the model in turn. e.g.:

$$\text{Pr(Conflict)} = f(\text{democracy, trade, power balance, contiguity, etc.})$$

$$\text{Pr(Voting)} = f(\text{education, income, party ID, race, gender, etc.})$$

- But in theory different objectives can require very different x's that enter the model. Prediction: all direct causes of y; causal inference on x_i: all x_j's that *confound* the relationship between x_i and y.



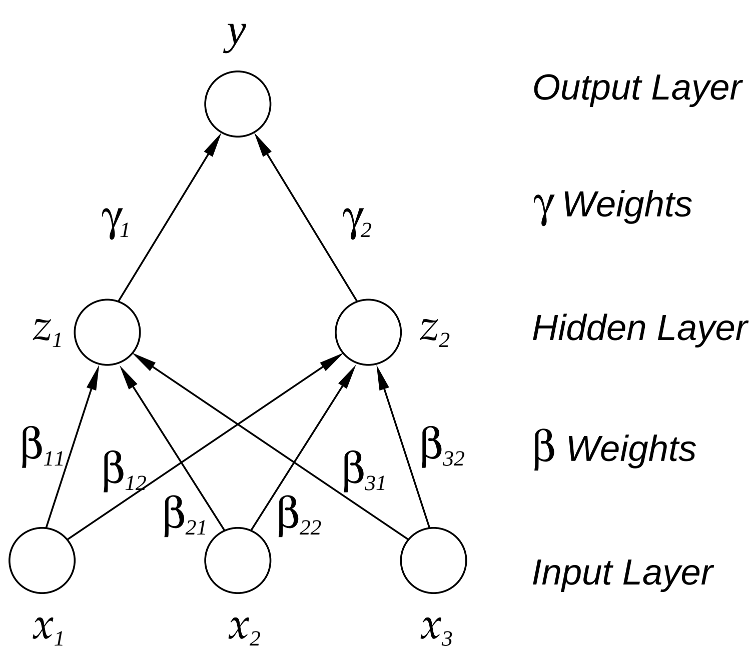
In this hypothetical causal structure:

- prediction of y: all x's;
- causal effect of x_1 on y: x_1 and x_2;
- causal effect of x_2 on y: x_2 (controlling for x_1, its consequence, leads to bias).
- causal effect of x_3 on y: x_3

- Finding the "right" set of control variables is the "Achilles heel of all social science inference because it is so very hard to do" (Brady 2002)
  - In practice, decision is often made "informally, on a case-by-case basis, resting on folklore and intuition rather than on hard mathematics." (Pearl 2000)
  - Thus different scholars studying the causal effects of the same x on the same y are often found to use different sets of control variables, guided by even slightly different substantive theories.
  - This can lead to not only changes in magnitude but often even reversal of signs in estimated key relationships across different studies, a phenomenon known as *Simpson's Paradox*.
- Zeng (2004) introduces new tools based on causal graph theory that address this and related issues, such as
  - the possibility of causal inference from observational data;
  - the "discovery" of underlying causal graphs from data;
  - graphical tools for control variable selection based on the causal graph.
  - improvement of non-parametric approaches for causal effect estimation.

Neural networks: Functional complexity in social science data

- Social science data generating function rarely, if ever, known; true functional forms likely complex
- Standard models such as linear regression, logit, probit, Tobit, Heckit, Poisson... all assume exact functional forms, and linearity one way or another.
  - Simplicity is an advantage only when the model is correct.
- In a series of papers, Zeng (1999, 2000b, 2004), King and Zeng (2001a), Beck, King, and Zeng (2000, 2004) introduce, adapt, apply, and extend a class of flexible, nonlinear models ("neural networks")
  - NN's are capable of approximating any unknown functional form to any degree of accuracy.
  - Our work reveals the rich nonlinear and interactive patterns in, for example, international conflict, state failure, and strategic voting data, which standard models are unable to handle by construction.


y
Output Layer
γ1
γ2
γ Weights
z1
z2
Hidden Layer
β11
β12
β21
β22
β31
β32
β Weights
Input Layer
x1
x2
x3

Social Networks: Modeling structural properties

- Social network analysis assists structural modeling by:
  - providing unambiguous, precise definitions of an array of graph theoretic measures capturing the properties of the whole system
  - allowing systematic modeling of endogenous dependence structure in random graphs representing statistical data.
- e.g., Modeling structural properties of IR data
  - little doubt that structural properties of the international system have important explanatory/predictive power
  - how to measure/model systematically? largely eluding previous research

Measuring structural characteristics: Actor degree and centrality

- Examining the N by N adjacency matrix X of the international system.
- Degree: in undirected graph, number of nodes adjacent to the actor i: $\Sigma_j X_{ij} = \Sigma_j X_{ji}$. In digraph, *in-degree* and *out-degree* are the numbers of nodes adjacent to and from respectively. Centrality: $\Sigma_j X_{ij}/N$. Actor with more ties with other actors more active and more visible.
  - e.g. Conflict involvement relation: a country with high degree/centrality is more conflict prone; Conflict initiation relation: higher degree indicates higher tendency of aggression and risk taking behavior.

    $\rightarrow$ Empirical test: conflict degree measures averaged over the past years in MID data powerful new input variable.
- Actor level measures. but of properties of actors *not* as isolated/independent entities, but embedded in the network–in this sense "structural" measures. (in IR, standard variables of state/dyad attributes do not have this feature.)

- Some other IR examples:
  - measuring “power” and “trade openness”. existing measures: “major power” indicator gross; “total trade as percentage of GDP” ignores other relevant aspects of openness (e.g., number of trade partners).

    → Use info from degree/centrality measures in pertinent networks (e.g. trade flows, cultural exchange, diplomatic relations, military interventions, conjoint treaty and international organization memberships) to refine measurement of “power”
  - degree/centrality in handling spatial dependence: e.g., spill over effects of democracy and conflict involvement. Ward and Gleditsch (N.D.)'s variables (number of neighboring countries in conflict; average level of democracy in proximate countries) in fact degree measures.

Measuring structural characteristics: Density and centralization

- system-level measures
- density: $\Sigma_{i,j} X_{ij}/N(N-1)$. For conflict data, a measure of rareness of events.
  - IR example: density measures on networks or subnetworks of such relations as trade, co-membership in international organizations, and neighboring a democratic state can provide informative measures, at the system level, of economic interdependence, growth of international norms and institutions, and degree of democratization. Current measures of these three key suppressors of violence largely restricted to the dyadic level (Russett and Oneal 2001, Russett 2003b).

- centralization: measures extent to which the system is centralized. One measure: the variance of the actor-level degrees or centrality scores. (variability and spread of actor-level centrality scores.)

  In a highly centralized system, some actors are much more central and important than others, who may be viewed as residing in the periphery of the system.

  - IR example: widely used measure of power structure in existing literature indicates whether the system is "bipolar" or "multi-polar," (whether there are two or more "great powers".) gross and insensitive to finer features. cold war era bipolar, most other times multi-polar.
  - centralization scores from pertinent international networks (such as those considered in constructing measures of "power") can enrich measurement of centralization structure of the international system.

Measuring structural characteristics: Cohesive subgroups

- One definition: cliques. Cohesive subgroups in the sense of having complete mutuality of ties among group members; Another: based on frequency of ties among members: k-*cores* of the graph. A k-core is a subgraph in which each node is adjacent to at least k other nodes in the subgraph.
  - e.g., cohesive subgroups in the network of treaty co-membership relevant to the study of coalition formation and alliance configuration

    (Nations in the same coalition are less likely to fight with each other — a generalization of the existing dyadic level "Ally" variable)
- Structural knowledge of the whole network: if the cohesive groups largely overlapping and contain most of the actors, the network as a whole is more cohesive, otherwise more fragmented.

Measuring structural characteristics: Roles and subgroup positions

- another approach to grouping actors in a network. Those in the same position are "equivalent". One definition is "structural equivalence". Two actors i and j are structurally equivalent if they have identical ties to and from all other actors on all relations under consideration: $X_{ikr} = X_{jkr}$ and $X_{kir} = X_{kjr}$ for any k in the node set and any r in the relation set.
- In practice unlikely to observe exactly equivalent actors, but can measure the degree they approach equivalence by comparing the similarity of the entries in X, based on measures of either correlation or distance.
- similarity measures can be used to partition the actors into groups/positions.
  - easy to see that both the "τ_b" (Bueno de Mesquita 1975) and the "S" scores (Signorino and Ritter 1999) widely used in the IR literature are in fact measures of structural equivalence in the graph of the alliance network.
  - Casting these measures into graph theory framework immediately suggests other possibilities based on alternative definitions of "equivalence."
- Preliminary results show that *all* measures discussed so far can be recovered from sub-sampled data from case-control designs that are more efficient for rare event data, provided N and n_0/N_0 known.

Modeling Endogenous Dependence Structure

- IR data motivation: Interdependence among actors and actions in international relations widely recognized:

> "Unlike, say, simple random survey sampling, dyadic observations in international conflict data have *complex dependence structure*. In a survey, observations 1 and 2 are two people who almost surely have never met and have no relationship. In contrast, in dyadic data, observation 1 may be U.S.-Iraq; observation 2, U.S.-Iran; and observation 3, Iraq-Iran. the dependence among these separate observations is complicated, central to our theories about the international system, critical for our methodological analysis, and ignored by most previous researchers." (King 2001)

- recent developments in models for random graphs provide useful tools for this task.

- random graph: graph with random edges, so that each element in X is a random variable. e.g., observed dyadic conflict data viewed as realizations of the underlying random graph.
  - Denote by $x = [x_{ij}]$ a realization of the graph X (i.e., an observed dataset). Dependences among elements of X mean that $Pr(X = x) \neq \Pi_{i \neq j} Pr(X_{ij} = x_{ij})$.
    * this condition rarely recognized in empirical models. The most widely used standard logit model, for example, implicitly assumes independence, i.e., $Pr(X = x) = \Pi_{i \neq j} Pr(X_{ij} = x_{ij})$.
    * need instead a general expression for $Pr(X = x)$ that allows any dependence structure among the elements of X, i.e., the dyads.
- a class of random graph models ("p*" models):

  $$Pr(X = x) = \frac{\exp(\theta' z(x))}{c(\theta)} \tag{1}$$

  $z(x)$ a vector of network statistics, *any* functions of elements of x; θ the vector of model parameters, and $c(\theta)$ a normalization factor. this allows general dependence.
- what enters $z(x)$?

  Hammersley-Clifford theorem: cliques (single nodes or complete subgraphs) of the dependence graph are the *sufficient subgraphs* for the representation of $Pr(X = x)$, so that $z(x)$ only needs to contain indicators of these cliques.

- the *Dependence graph*, D, of a graph G depicts the dependence structure among dyads of G: all possible dyads of G as nodes, tie between two dyads if they are conditionally dependent given the remaining dyads.
  - Different patterns of interdependence among the dyads lead to different connection patterns in the dependence graph.
  - if all pairs of dyads are conditionally independent (usual assumption), the dependence graph would be empty and have no ties.
  - most commonly assumed dependence structure: *Markov dependence*. dyads sharing at least one member are conditionally dependent.
    * e.g., U.S.-Iraq and U.S.-Iran; U.S.-Iraq and Iraq-Iran, and U.S.-Iran and Iraq-Iran.
    * can show that for Markov dependence, the cliques of D are just the *triangles* and *k-stars*, $k = 1, 2, \ldots, N-1$, of G.
      · E.g., U.S.-Iraq: 1-star; U.S.-Iraq and U.S.-Iran form a 2-star.
    * So by the Hammersley-Clifford theorem, pertinent network statistics are just the number of triangles and k-stars (which can be replaced by number of nodes in G that are of degree k).

- Estimation: Standard llik techniques difficult to apply due to the normalization term. do pseudo-likelihood based on conditional probabilities.

  Denote X_{ij}^+ the graph with the edge ij forced to be present, X_{ij}^- the graph with the edge ij forced to be absent, and X_{ij}^C the graph with the edge ij "missing". Then:

$$\begin{aligned} Pr(X_{ij} = 1 | X_{ij}^C) &= \frac{Pr(X_{ij}^+)}{Pr(X_{ij}^+) + Pr(X_{ij}^-)} \\ &= \frac{\exp(\theta' z(x_{ij}^+))}{\exp(\theta' z(x_{ij}^+)) + \exp(\theta' z(x_{ij}^-\ ))} \\ &= \frac{1}{1 + \exp(-\theta'(z(x_{ij}^+) - z(x_{ij}^-)))} \qquad (2) \end{aligned}$$

  Expression (2) identical to a logit model probability with $z(x_{ij}^+) - z(x_{ij}^-) = \delta(x_{ij})$ as the explanatory variables. $\delta(x_{ij})$: differences in network statistics when the edge ij changes from absent to present

  $\rightarrow$ can be estimated with logit routines.

Empirical results

- Data: militarized international dispute (MID) data set, from Beck, King and Zeng (2000). 23,529 interstate dyad-years between 1947 and 1989, 976 (4.1%) MIDs (1's) and the rest 0's.
- Using the logit model in BKZ as the "standard logit" model for comparison. (logit as first cut). Explanatory variables: "Contiguity", "Ally", "Similarity" (similarity in the dyad's foreign policy portfolios), "Asymmetry" (balance of of power), "Dema" and "Demb" (using min(dema,demb) here), "Peace Years".
- Conjecture 1: properties of individual states/dyads embedded in the international network, in addition to attributes of them as independent entities, are important.
  - In particular, examine past conflict tendencies of the dyad as members of the international community, see how that influence current conflict behavior within the dyad.
  - measure: *degree* of the country in the conflict graph for the international network, averaged over the past years. then take min(conf1,conf2).
  - results: positive, highly significant coefficient. increase llik -3178.65 to -3066.48. doubles improvement of test data performance by standard model over base. Marginal effects plot reveals much stronger effects than any other standard variables.

- Conjecture 2: Markov dependence
  - meaning of transitivity in conflict relation unclear; focus on the k-stars in the test.
  - results confirm dependence. network statistics improve test set ROC area over base model 4 times as much as standard logit.
  - negative sign of coeffs: graphs with less clustering more likely to realize; higher order clustering less likely than lower level clustering.

    $\rightarrow$ consistent with data: very rare events, and even rarer cases where one country involved in multiple conflicts at the same time.

Table 1: Models of International Conflict

Variables	P.Y.-Base Coeff	P.Y.-Base P-value	Standard Logit Coeff	Standard Logit P-value	+Past Degree Coeff	+Past Degree P-value	+Network Stats Coeff	+Network Stats P-value
Peace Years	-.181	0.000	-.153	0.000	-.138	0.000	-.135	0.000
Contiguity	—	—	1.046	0.000	1.179	0.000	1.411	0.000
Ally	—	—	-.325	0.001	-.227	0.020	-.205	0.037
Similarity	—	—	-.282	0.000	-.084	0.336	-.006	0.949
Asymmetry	—	—	-.882	0.000	-.329	0.017	-.657	0.000
Min-Dem.	—	—	-.045	0.000	-.038	0.000	-.036	0.000
Min-Conf.	—	—	—	—	.868	0.000	.5212	0.000
D1star	—	—	—	—	—	—	-.675	0.000
D2star	—	—	—	—	—	—	-.903	0.000
D3star	—	—	—	—	—	—	-.910	0.000
D4star	—	—	—	—	—	—	-.434	0.002
D5star	—	—	—	—	—	—	-.472	0.000
Constant	-1.619	0.000	-1.786	0.000	-2.698	0.000	-1.85	0.000
N	23255		23255		23255		23255	
Log-likelihood	-3379.83		-3178.65		-3066.48		-2990.036	
Out-of-sample ROC	.826		.834		.843		.859	
ROC gain over base	0		.008		.017		.033	

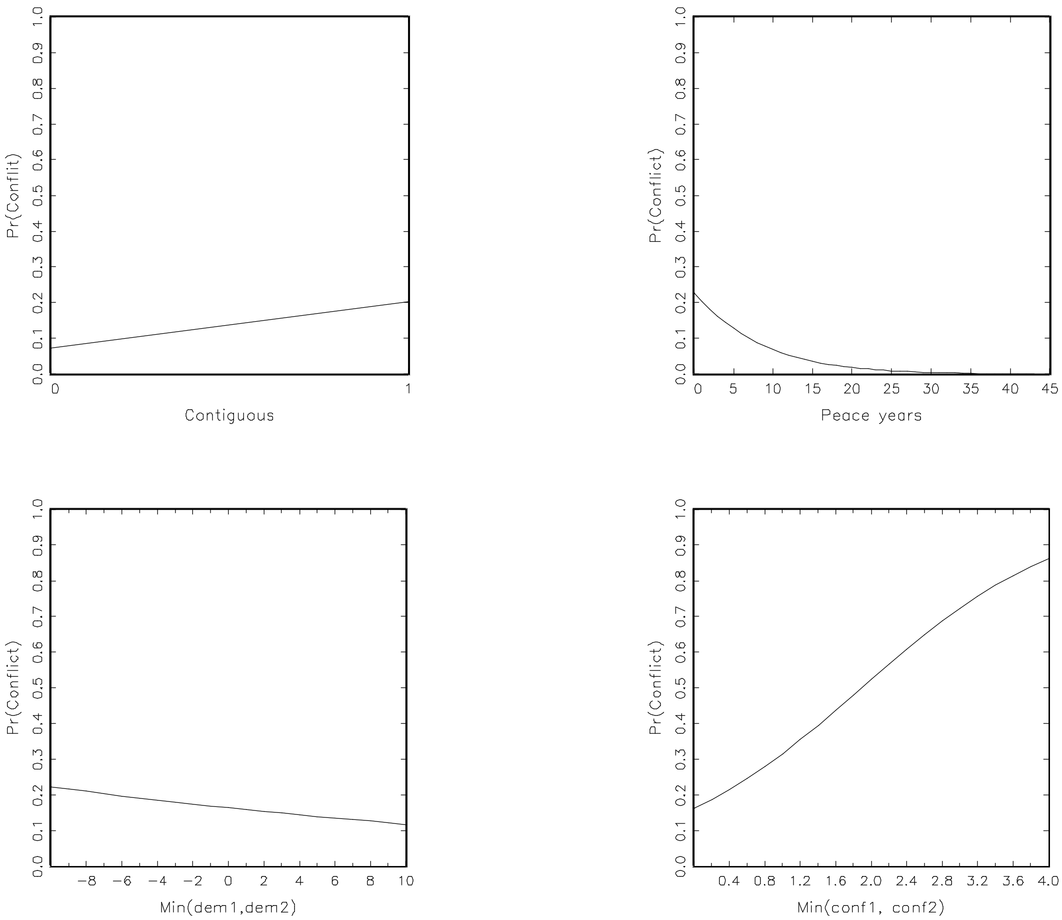


Figure 1: *Marginal effects of input variables in the MID model.*

Conclusions

- Graphical methods and models are useful tools for improving prediction and causal inference in empirical social science.
- Social network analysis improves structural modeling both through providing graph theoretic measures of structural characteristics of the system, and by allowing systematic modeling of endogenous dependence structures in random graphs representing statistical data.
- Empirical example on international conflict data demonstrates the usefulness of the methods.
- Future work would explore extensions of the methods to accommodate special features of social science data, such as functional complexity and rareness of certain events, in addition to applications to other substantively important problems.