Opportunity and Choice in Social Networks

Paolo Pin, Silvio Franz and Matteo Marsili

NOTA DI LAVORO 24.2008

MARCH 2008

CTN – Coalition Theory Network

Paolo Pin, *The Abdus Salam International Centre for Theoretical Physics, Dipartimento di Scienze Economiche, Università Ca' Foscari di Venezia*
Silvio Franz, *The Abdus Salam International Centre for Theoretical Physics, Centre Scientifique d'Orsay, Universite Paris-Sud 11*
Matteo Marsili, *The Abdus Salam International Centre for Theoretical Physics*

Opportunity and Choice in Social Networks

Summary

Our societies are heterogeneous in many dimensions such as census, education, religion, ethnic and cultural composition. The links between individuals - e.g. by friendship, marriage or collaboration - are not evenly distributed, but rather tend to be concentrated within the same group. This phenomenon, called imbreeding homophily, has been related to either (social) preference for links with own--type individuals (choice-based homophily) or to the prevalence of individuals of her same type in the choice set of an individual (opportunity-based homophily). We propose an indicator to distinguish between these effects for minority groups. This is based on the observation that, in environments with unbiased opportunities, as the relative size of the minority gets small, individuals of the minority rarely meet and have the chance to establish links together. Therefore the effect of choice--based homophily gets weaker and weaker as the size of the minority shrinks. We test this idea across the dimensions of race and education on data on US marriages, and across race on friendships in US schools, and find that: for what concerns education i) opportunity--based homophily is much stronger than choice--based homophily and ii) they are both remarkably stationary in time; concerning race iii) school friendships do not exhibit opportunity-based homophily, while marriages do, iv) choice-based homophily is much stronger for marriages than for friendships and v) these effects vary widely across race.

Keywords: Social Networks, Choice-Based and Opportunity-Based Homophily

JEL Classification: D85, J11, J12

This paper was presented at the 13th Coalition Theory Network Workshop organised by the Fondazione Eni Enrico Mattei (FEEM), held in Venice, Italy on 24-25 January 2008.
Some results in this paper have been circulated in an unpublished note titled “New Conditions for Core Existence in Coalitional NTU Games” (2001). I would like to thank Herbert Scarf and Donald Smythe for their valuable comments on my earlier work. All errors, of course, are my own.

Address for correspondence:

Paolo Pin
Dipartimento di Scienze Economiche
Università Ca' Foscari di Venezia
Dorsoduro 3246
30123 Venezia
Italy
E-mail: pin@unive.it

Opportunity and choice in social networks

Paolo Pin,[1,2] Silvio Franz,[1,3] and Matteo Marsili[1]

[1] *The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy*
[2] *Dipartimento di Scienze Economiche, Università Ca' Foscari di Venezia, Italy*
[3] *Centre scientifique d'Orsay, Université Paris-Sud 11, France*

(Dated: February 11, 2008)

Our societies are heterogeneous in many dimensions such as census, education, religion, ethnic and cultural composition. The links between individuals – e.g. by friendship, marriage or collaboration – are not evenly distributed, but rather tend to be concentrated within the same group. This phenomenon, called *imbreeding homophily*, has been related to either (social) preference for links with own–type individuals (*choice–based* homophily) or to the prevalence of individuals of her same type in the choice set of an individual (*opportunity–based* homophily). We propose an indicator to distinguish between these effects for minority groups. This is based on the observation that, in environments with unbiased opportunities, as the relative size of the minority gets small, individuals of the minority rarely meet and have the chance to establish links together. Therefore the effect of choice–based homophily gets weaker and weaker as the size of the minority shrinks. We test this idea across the dimensions of race and education on data on US marriages, and across race on friendships in US schools, and find that: for what concerns education i) opportunity–based homophily is much stronger than choice–based homophily and ii) they are both remarkably stationary in time; concerning race iii) school friendships do not exhibit opportunity-based homophily, while marriages do, iv) choice-based homophily is much stronger for marriages than for friendships and v) these effects vary widely across race.

Integration is a major concern of our societies, whose relevance has increased as an effect of globalization. The prevalence of relations between individuals of the same type or community over links across types – a well known phenomenon called (inbreeding) *homophily* in sociology [1–8] – has been related to either *opportunity–based* or *choice–based* homophily [5, 7]: while the former (also called *induced* homophily) refers to a bias towards same–type neighbors in the underlying social network, the latter reflects a bias towards same–type links in the *collective* choice of mutual relations, among those possible in a given neighborhood of the social network. Here, choice–based bias depends on the choices of the population as a whole, and cannot be related to choices of the individual. For example, Ref. [9] shows that individual choices influence in non–trivial ways the aggregate outcome and Ref. [10] argues that biased mixing of a minority may be due to homophily of both majority and minority individuals.

Also, the opportunities which an individual faces when choosing whom to establish a relation with may well be shaped by past choices of that individual and others. For example, T.C. Schelling has vividly shown that even very weak preferences for homophilous relations in residential choice, can lead to strong spacial segregation [11, 12]. However, even if choices influence opportunities across time, it is meaningful to make a distinction between choices and opportunities the individuals face at a given time.

There are many dimensions (ethnical, religion, education, age, census etc.) which are likely to influence, to different degrees, the formation of links between individuals, and these are correlated in complex ways. Disentangling their effect is a non-trivial task [13, 14]. Still it makes sense to discuss biases in the choices and opportunities along a single dimension, provided due caution is paid in drawing causal relations. This is particularly true in cases where the pattern of interaction is shaped by institutions. For example, friendships between school students is a matter of individual choice but their pattern of interaction is largely shaped by institutions (clubs, sport teams, academic tracking [15], etc). So while it is natural to expect choice–based homophily, the presence of opportunity–based homophily may be a matter of concern for policy makers.

Our aim, in the present work, is to provide an indicator which separates the effects of opportunity–based homophily (OBH) and of choice–based homophily (CBH) for a minority group inside a larger population. The idea is that, in the absence of opportunity biases, CBH has an effect which is proportional to the size of the minority, when the latter is small, simply because individuals of that minority have no opportunities to meet. Therefore, an excess of inter-type links in very small minorities must be due to OBH.

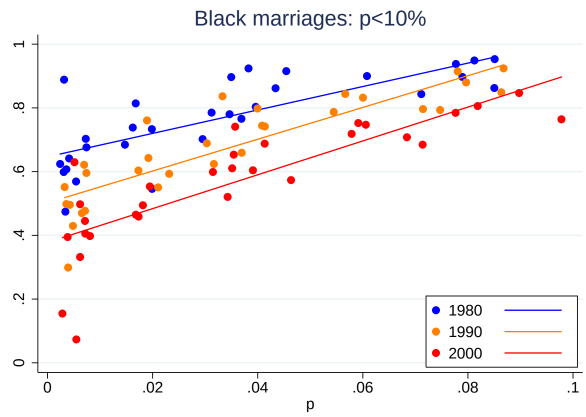
To be more precise, let q be the ratio of same–type links for a member of a minority and let p be the relative size (frequency) of this minority. Following Coleman [17], Inbreeding Homophily can be measured by the index

$$H \equiv \frac{q-p}{1-p} \ , \tag{1}$$

which is the excess of minority type links normalized so that baseline homophily ($q = p$) [2] corresponds to $H = 0$ and complete segregation ($q = 1$) yields $H = 1$. $H(p)$ depends non–linearly on p [3, 8, 9], but for small p, we can assume [18]

$$H(p) \simeq A + Bp \ . \tag{2}$$

The above observation implies that we should have $A = 0$ in a population of homophilous individuals, with no OBH. Therefore, A can be taken as an index of OBH. This general observation can be detailed in a simple probabilistic model, which explicitly takes into account the two effects (see Model). OBH is modeled by the frequency π of the minority in the typical neighborhood of the social network of a minority individual, for $p \to 0$. The actual social relations are chosen on the social network thus defined, with a same–type link of minority individuals being chosen x times more likely than a different–type link. π and x, in the model, can be derived from A and B.



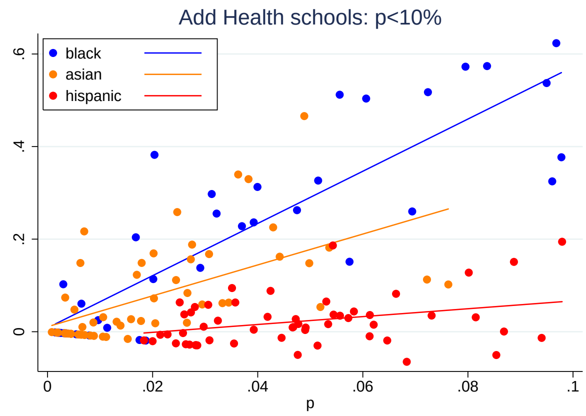


FIG. 1: IPUMS data for the Black minority (above) are based on same–race marriages in American States [19]. Each point represents the Black minority in one State in one of the three surveys (1980, 1990, 2000). On the x–axis we have the percentage of the minority (the maximum threshold is 10%). On the y–axis the Imbreeding Homophily measure defined in (1). Lines represent linear fits for each survey. Add Health data (below) are based on same–race friendships in American schools [20]. Each point represents a minority in one school. Linear fits are made for each minority.

IPUMS marriages in US states

no threshold	**n**	**B**	$\pm$	**A**	$\pm$	**x**	$\pm$	π	$\pm$
Top Educ. 1980	51	0.7	0.4	0.22	0.04	1.2	0.7	0.113	0.04
Top Educ. 1990	51	1.0	0.3	0.20	0.04	1.5	0.5	0.089	0.02
Top Educ. 2000	51	1.1	0.3	0.19	0.04	1.7	0.4	0.081	0.02
10% **threshold**									
Black 1980	30	3.7	1.1	0.65	0.06	29.4	13.4	0.057	0.02
Black 1990	30	5.0	1.0	0.50	0.04	20.1	5.4	0.046	0.01
Black 2000	30	5.3	1.4	0.38	0.06	13.7	4.3	0.040	0.01
Asian 1980	50	8.3	4.0	0.38	0.06	21.9	11.3	0.026	0.01
Asian 1990	50	4.8	2.3	0.45	0.06	15.6	8.3	0.046	0.02
Asian 2000	49	5.5	2.2	0.47	0.06	19.1	8.7	0.042	0.02
Native 1980	50	6.8	1.6	0.16	0.04	9.6	2.5	0.018	0.01
Native 1990	50	4.5	1.3	0.14	0.04	6.2	1.8	0.023	0.01
Native 2000	49	4.1	1.3	0.17	0.04	5.9	2.0	0.029	0.01
Hispanic 1980	50	6.5	2.7	0.41	0.06	18.4	8.4	0.034	0.01
Add Health friendships in US schools									
Black	39	5.6	1.0	0.01	0.39	5.7	4.7	0.001	0.06
Asian	56	3.3	1.3	0.01	0.04	3.4	1.3	0.002	0.01
Hispanic	55	0.9	0.7	-0.02	0.04	0.8	0.6	-0.010	0.02

TABLE I: Every line represents a minority in one survey. For the first three the minority *Top Educ.* represents all those people who have spent at least 4 years in college. For the remaining lines the minority is represented by a race. n is the number of observations. We compute A, B and their 95% confidence interval, with a linear regression of p versus H. In the cases concerning education there is no threshold on p. (the reason for this is that p has almost doubled in every State between 1980 and 2000). For the remaining regressions we take only those p below 10% (results are qualitatively robust to a change of this threshold). We compute x, π, and their relative 95% confidence interval, with the model described in the Model.

We illustrate this idea on empirical data on marriages [19] and friendships [20], where individuals are identified by race and (in marriages) by the level of education attained. The datasets pertain to different environments – single American States for marriages [19] or single schools for friendships [20] – with different relative share p of minority individuals. For each environment, we measure the inbreeding homophily index H and compute A and B from a linear fit for small p, π and x are computed from the model (see Model). Fig. 1 shows a sample of the results which are collected in Table I. The top panel shows the fit of $H(p)$ for marriages in the minority of Blacks, in three subsequent surveys (1980, 1990 and 2000). First, we observe strong CBH (x), which is also observed for other types (see Table I). Also notice that $A > 0$, indicating OBH ($\pi > 0$), though this effect has been declining over time. OBH in marriages is observed also for other minorities, but it is close to negligible for Native Americans, and its time dependence is much weaker. On the contrary, CBH and OBH are remarkably stable across time with respect to education, to the point that data appears to lie on the same master line, though p values have increased of 46% from 1980 to 2000.

In school friendships (bottom panel), choice–based homophily is still high and significative, but much less than for the marriages considered above. Also here x strongly varies from race to race. It is moreover acceptable, for all the regressions and all the races, to assume that A, and hence π is equal to 0, implying no OBH. This is what we would assume from an environment like a school, where the class formation should be independent on races [15].

Summarizing, we have proposed a method to disentangle choice–based from opportunity–based sources of homophily (CBH and OBH respectively). Our case study on two data sets shows that, for what concerns marriages alone: (i) OBH is stronger for *top educated* people than for any racial minority, but CBH is much weaker. (ii) Looking at different time windows, for marriages, there is a clear decrease of both measures of homophily for Blacks between 1980, 1990 and 2000. This time–dependence is not so evident for the other races and especially not for *top educated* people. For what concerns the racial dimension: (iii) School friendships do not exhibit OBH (compared to the school population), while marriages do. (iv) CBH is much stronger for marriages than for friendships. (v) The values of both are strictly race–dependent: Blacks exhibit the strongest CBH and (in marriages, if compared to the population of the American States) OBH; Hispanics exhibit the lowest values of both (which could be both accepted as uninfluent in the school data).

There are several interesting extensions of our analysis to other dimensions such as religion or census. The outcome of our analysis needs to be critically evaluated, as our distinction between choice and external constraints is theoretical at best. If anything, it may help in identifying those institutional constraints which hamper fruitful exchanges between members of our society.

[1] Lazarsfeld, P.F, & Merton, R.K. Friendship as a social process: a substantive and methodological analysis. In Berger, M. (ed.) *Freedom and Control in Modern Society* 18–66, New York: Van Nostrand (1954).

[2] Blau, P.M. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press (1977).

[3] Marsden, P.V. Models and methods for characterizing the structural parameters of groups. *Social Networks* **3**, 1–27 (1981).

[4] Rytina, S. & Morgan, D.L. The Arithmetic of Social Relations: The Interplay of Category and Network. *The American Journal of Sociology* **88**(1), 88–113 (1982).

[5] McPherson, M. & Smith-Lovin, L. Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups. *American Sociological Review* **52**(3), 370–379 (1987).

[6] Marsden, P.V. Homogeneity in confiding relations. *Social Networks* **10**, 57–76 (1988).

[7] McPherson, M., Smith-Lovin, L. & Cook, J.M. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* **27**, 415–444 (2001).

[8] Moody, J. Race, school integration, and friendship segregation in America. *American Journal of Sociology* **107**(3), 679–716 (2001).

[9] Currarini, S., Jackson, M.O., & Pin, P. An Economic Model of Friendship: Homophily, Minorities and Segregation. Mimeo (2007).

[10] Jego, C. & Roehner, B.M. A physicist's view of the notion of 'racism'. Mimeo (2007).

[11] Schelling, T. Dynamic models of segregation. *Journal of Mathematical Sociology* **1**, 143–86 (1971).

[12] Vinkovič, D. & Kirman, A. A physical analogue of the Schelling model. *Proceedings of the National Academy of Sciences* **103**(51), 19261–19265 (2006).

[13] Bisin, A. & Topa, G. Empirical Models of Cultural Transmission. *Journal of the European Economic Association* **1**, 363–375 (2003).

[14] Bisin, A., Topa, G. & Verdier, G. Religious Intermarriage and Socialization in the United States. *Journal of Political Economy* **112**(3), 615-664 (2004).

[15] One of the criteria for class formation in the U.S. is *academic tracking*. An extreme example is: best–grades students in one class, and so on to worst–grades students in the last class. If school performances were correlated with race, then academic tracking could lead to a race bias. Past literature[16] and Add Health data show that this bias, if present, is ininfluent.

[16] Gamoran, A. & Mare, R.D. Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality? *American Journal of Sociology* **94**(5), 1146–1183 (1989).

[17] Coleman, J. Relational analysis: the study of social organizations with survey methods. *Human Organization* **17**, 28–36 (1958).

[18] In some cases we found data to be better described by a non-linear relation $H(p) = \alpha + \beta p^\gamma + \ldots$ with $\gamma \approx 0.5$. However, while the qualitative nature of the conclusions discussed here does not change, it is hard to justify a non-linear regression or the specific value of $\gamma = 1/2$. This is why this issue will not be discussed further here.

[19] The Integrated Public Use Microdata Series (IPUMS). are data from surveys conducted in the 51 American States by the Minnesota Population Center, weighting households by their representativeness (http://usa.ipums.org). We are using the 5% population samples from years 1980, 1990 and 2000. Even if the classification is much more accurate we will consider four main racial types of minorities: Black, Asian, Native American and Hispanic (present only for the 1980 survey).

[20] The National Longitudinal Study of Adolescent Health (Add Health) are data from surveys conducted in 1994 in a sample of 84 American high schools and middle schools, by the UNC Carolina Population Center (http://www.cpc.unc.edu/addhealth). Students were self–reporting their race (we consider: Black, Asian and Hispanic) and could nominate their friends from a list of all the other students in their school. We consider a link whenever at least one of the two students nominate the other one. Results are robust also to only reciprocal links, but in this case many data are missing and variance increases.

[21] McFadden, D. Conditional logit analysis of qualitative choice behavior. In Zarembka, P. (ed.) *Frontiers in Econometrics* 105–142, New York: Academic Press (1974).

APPENDIX: THE MODEL

We imagine a society whose individuals are ex-ante divided in different types, whose number N is fixed and large. Let p be the fraction of a particular minority in the population. The local environment of each individual is defined by an underlying social network, with K links for each individuals. K is supposed to represent the number of *possible* links from which the actual relations (marriage, friendship, etc) of a particular individual are drawn. We assume that K is (much) larger than the actual number k of relations each individual establishes, but much smaller than N (for schools, $k \sim 6$ whereas $K \sim 30$ may be taken as the typical class size, and N is in the order of hundreds). Individuals are distributed inhomogeneously on the social network, in such a way that the average frequency of the minority in the neighborhood of a minority individual is

$$\bar{p}(p) = \pi + (1-\pi)p \quad , \tag{A.1}$$

with $\pi \in (0,1)$. The relation is taken to be linear for simplicity, with $\bar{p}(1) = 1$.

We assume each individual of the minority has k links, and we assign them in the following way: *i)* choose an individual of the minority at random, *ii)* if he still has an unassigned link, choose one of the unassigned links in her neighborhood with a statistical weight $1+x$ times larger for links to minority individuals than to majority ones; *iii)* stop when all links of the minority are assigned. For marriages we consider a bipartite network in which all neighbors of an individual are of the opposite sex. x has a naïve interpretation in term of utility in discrete choice models [21], but it also reflects more complex aspects of the matching problem (see e.g. [9]).

On average, each individual will have

$$\begin{cases} k\frac{\bar{p}(1+x)}{(1+x)\bar{p}+(1-\bar{p})} & \text{same–type links,} \\ k\frac{1-\bar{p}}{(1+x)\bar{p}+(1-\bar{p})} & \text{different–type links.} \end{cases} \tag{A.2}$$

Therefore the ratio q of same–type links, in the whole population, over all links is $q \simeq \frac{\bar{p}(1+x)}{1+x\bar{p}}$, and this for small p leads to

$$H(p) = \pi\frac{1+x}{1+\pi x} + x\left(\frac{1-\pi}{1+\pi x}\right)^2 p + \mathcal{O}(p^2) \tag{A.3}$$

from which we can read the values of A and B in Eq. (2). Likewise, from A and B we can infer that

$$x \;\simeq\; \frac{B}{(1-A)^2} \;\;,\;\; \pi \;\simeq\; \frac{A(1-A)}{1-A+B} \quad . \tag{A.4}$$

As a check, we generated synthetic data sets using the model, with x and π given in Table I, and performed a linear regression of $H(p)$. The resulting values of A' and B' were found to be within the 95% confidence intervals reported in Table I for A and B, in almost all cases. We attribute the discrepancy to a systematic bias due to non–linear terms in $H(p)$ [18]. These effects are particularly strong when x is large. These issues would require a more sophisticated estimation techniques, which goes beyond the scope of the present paper.

NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI

Fondazione Eni Enrico Mattei Working Paper Series

Our Note di Lavoro are available on the Internet at the following addresses:

http://www.feem.it/Feem/Pub/Publications/WPapers/default.htm
http://www.ssrn.com/link/feem.html
http://www.repec.org
http://agecon.lib.umn.edu
http://www.bepress.com/feem/

NOTE DI LAVORO PUBLISHED IN 2008

CCMP	1.2008	*Valentina Bosetti, Carlo Carraro and Emanuele Massetti*: Banking Permits: Economic Efficiency and Distributional Effects
CCMP	2.2008	*Ruslana Palatnik and Mordechai Shechter*: Can Climate Change Mitigation Policy Benefit the Israeli Economy? A Computable General Equilibrium Analysis
KTHC	3.2008	*Lorenzo Casaburi, Valeria Gattai and G. Alfredo Minerva*: Firms' International Status and Heterogeneity in Performance: Evidence From Italy
KTHC	4.2008	*Fabio Sabatini*: Does Social Capital Mitigate Precariousness?
SIEV	5.2008	*Wisdom Akpalu*: On the Economics of Rational Self-Medication
CCMP	6.2008	*Carlo Carraro and Alessandra Sgobbi*: Climate Change Impacts and Adaptation Strategies In Italy. An Economic Assessment
ETA	7.2008	*Elodie Rouvière and Raphaël Soubeyran*: Collective Reputation, Entry and Minimum Quality Standard
IEM	8.2008	*Cristina Cattaneo, Matteo Manera and Elisa Scarpa*: Industrial Coal Demand in China: A Provincial Analysis
IEM	9.2008	*Massimiliano Serati, Matteo Manera and Michele Plotegher*: Econometric Models for Electricity Prices: A Critical Survey
CCMP	10.2008	*Bob van der Zwaan and Reyer Gerlagh*: The Economics of Geological CO_2 Storage and Leakage
KTHC	11.2008	*Maria Francesca Cracolici and Teodora Erika Uberti*: Geographical Distribution of Crime in Italian Provinces: A Spatial Econometric Analysis
KTHC	12.2008	*Victor Ginsburgh, Shlomo Weber and Sheila Weyers*: Economics of Literary Translation. A Simple Theory and Evidence
NRM	13.2008	*Carlo Giupponi, Jaroslav Mysiak and Alessandra Sgobbi*: Participatory Modelling and Decision Support for Natural Resources Management in Climate Change Research
NRM	14.2008	*Yaella Depietri and Carlo Giupponi*: Science-Policy Communication for Improved Water Resources Management: Contributions of the Nostrum-DSS Project
CCMP	15.2008	*Valentina Bosetti, Alexander Golub, Anil Markandya, Emanuele Massetti and Massimo Tavoni*: Abatement Cost Uncertainty and Policy Instrument Selection under a Stringent Climate Policy. A Dynamic Analysis
KTHC	16.2008	*Francesco D'Amuri, Gianmarco I.P. Ottaviano and Giovanni Peri*: The Labor Market Impact of Immigration in Western Germany in the 1990's
KTHC	17.2008	*Jean Gabszewicz, Victor Ginsburgh and Shlomo Weber*: Bilingualism and Communicative Benefits
CCMP	18.2008	*Benno Torgler, María A.GarcíaValiñas and Alison Macintyre*: Differences in Preferences Towards the Environment: The Impact of a Gender, Age and Parental Effect
PRCG	19.2008	*Gian Luigi Albano and Berardino Cesi*: Past Performance Evaluation in Repeated Procurement: A Simple Model of Handicapping
CTN	20.2008	*Pedro Pintassilgo, Michael Finus, Marko Lindroos and Gordon Munro* (lxxxiv): Stability and Success of Regional Fisheries Management Organizations
CTN	21.2008	*Hubert Kempf and Leopold von Thadden* (lxxxiv): On Policy Interactions Among Nations: When Do Cooperation and Commitment Matter?
CTN	22.2008	*Markus Kinateder* (lxxxiv): Repeated Games Played in a Network
CTN	23.2008	*Taiji Furusawa and Hideo Konishi* (lxxxiv): Contributing or Free-Riding? A Theory of Endogenous Lobby Formation
CTN	24.2008	*Paolo Pin, Silvio Franz and Matteo Marsili* (lxxxiv): Opportunity and Choice in Social Networks

(lxxxiv) This paper was presented at the 13th Coalition Theory Network Workshop organised by the Fondazione Eni Enrico Mattei (FEEM), held in Venice, Italy on 24-25 January 2008.