

1

The Social Structure of Competition

A player brings capital to the competitive arena and walks away with profit determined by the rate of return where the capital was invested. The market production equation predicts profit: invested capital, multiplied by the going rate of return, equals the profit to be expected from the investment. You invest a million dollars. The going rate of return is 10 percent. The profit is one hundred thousand dollars. Investments create an ability to produce a competitive product. For example, capital is invested to build and operate a factory. Rate of return is an opportunity to profit from the investment.

The rate of return is keyed to the social structure of the competitive arena and is the focus here. Each player has a network of contacts in the arena. Something about the structure of the player's network and the location of the player's contacts in the social structure of the arena provides a competitive advantage in getting higher rates of return on investment. This chapter is about that advantage. It is a description of the way in which social structure renders competition imperfect by creating entrepreneurial opportunities for certain players and not for others.¹

Opportunity and Capital

A player brings at least three kinds of capital to the competitive arena. Other distinctions can be made, but three are sufficient here. First, the player has financial capital: cash in hand, reserves in the bank, investments coming due, lines of credit. Second, the player has human capital. Your natural qualities—charm, health, intelligence, and looks—combined with the skills you have acquired in formal education and job experience give you abilities to excel at certain tasks.

Third, the player has social capital: relationships with other players.

You have friends, colleagues, and more general contacts through whom you receive opportunities to use your financial and human capital. I refer to opportunities in a broad sense, but I certainly mean to include the obvious examples of job promotions, participation in significant projects, influential access to important decisions, and so on. The social capital of people aggregates into the social capital of organizations. In a firm providing services—for example, advertising, brokerage, or consulting—there are people valued for their ability to deliver a quality product. Then there are “rainmakers,” valued for their ability to deliver clients. Those who deliver the product do the work, and the rainmakers make it possible for all to profit from the work. The former represent the financial and human capital of the firm. The latter represent its social capital. More generally, property and human assets define the firm’s production capabilities. Relations within and beyond the firm are social capital.

DISTINGUISHING SOCIAL CAPITAL

Financial and human capital are distinct in two ways from social capital. First, they are the property of individuals. They are owned in whole or in part by a single individual defined in law as capable of ownership, typically a person or corporation. Second, they concern the investment term in the market production equation. Whether held by a person or the fictive person of a firm, financial and human capital gets invested to create production capabilities. Investments in supplies, facilities, and people serve to build and operate a factory. Investments of money, time, and energy produce a skilled manager. Financial capital is needed for raw materials and production facilities. Human capital is needed to craft the raw materials into a competitive product.

Social capital is different on both counts. First, it is a thing owned jointly by the parties to a relationship. No one player has exclusive ownership rights to social capital. If you or your partner in a relationship withdraws, the connection, with whatever social capital it contained, dissolves. If a firm treats a cluster of customers poorly and they leave, the social capital represented by the firm-cluster relationship is lost. Second, social capital concerns rate of return in the market production equation. Through relations with colleagues, friends, and clients come the opportunities to transform financial and human capital into profit.

Social capital is the final arbiter of competitive success. The capital invested to bring your organization to the point of producing a superb product is as rewarding as the opportunities to sell the product at a profit. The investment to make you a skilled manager is as valuable as the

opportunities—the leadership positions—you get to apply your managerial skills. The investment to make you a skilled scientist with state-of-the-art research facilities is as valuable as the opportunities—the projects—you get to apply those skills and facilities.

More accurately, social capital is as important as competition is imperfect and investment capital is abundant. Under perfect competition, social capital is a constant in the production equation. There is a single rate of return because capital moves freely from low-yield to high-yield investments until rates of return are homogeneous across alternative investments. When competition is imperfect, capital is less mobile and plays a more complex role in the production equation. There are financial, social, and legal impediments to moving cash between investments. There are impediments to reallocating human capital, both in terms of changing the people to whom you have a commitment and in terms of replacing them with new people. Rate of return depends on the relations in which capital is invested. Social capital is a critical variable. This is all the more true when financial and human capital are abundant—which in essence reduces the investment term in the production equation to an unproblematic constant.

These conditions are generic to the competitive arena, which makes social capital a factor as routinely critical as financial and human capital. Competition is never perfect. The rules of trade are ambiguous in the aggregate and everywhere negotiable in the particular. The allocation of opportunities is rarely made with respect to a single dimension of abilities needed for a task. Within an acceptable range of needed abilities, there are many people with financial and human capital comparable to your own. Whatever you bring to a production task, there are other people who could do the same job—perhaps not as well in every detail, but probably as well within the tolerances of the people for whom the job is done. Criteria other than financial and human capital are used to narrow the pool down to the individual who gets the opportunity. Those other criteria are social capital. New life is given to the proverb that says success is determined less by what you know than by whom you know. As a senior colleague once remarked (and Cole, 1992: chaps. 7–8, makes into an intriguing research program), “Publishing high-quality work is important for getting university resources, but friends are essential.” Of those who are equally qualified, only a select few get the most rewarding opportunities. Of the products that are of comparably high quality, only some come to dominate their markets. The question is how.

WHO AND HOW

The competitive arena has a social structure: players trusting certain others, obligated to support certain others, dependent on exchange with certain others, and so on. Against this backdrop, each player has a network of contacts—everyone the player now knows, everyone the player has ever known, and all the people who know the player even though he or she doesn't know them. Something about the structure of the player's network and the location of the player's contacts in the social structure of the arena provides a competitive advantage in getting higher rates of return on investment.

Who

There are two routes into the social capital question. The first describes a network as your access to people with specific resources, which creates a correlation between theirs and yours. This idea has circulated as power, prestige, social resources, and more recently, social capital. Nan Lin and his colleagues provide an exemplar of this line of work, showing how the occupational prestige of a person's job is contingent on the occupational prestige of a personal contact leading to the job (Lin, 1982; Lin, Ensel, and Vaughn, 1981; Lin and Dumin, 1986). Related empirical results appear in Campbell, Marsden, and Hurlbert (1986), De Graaf and Flap (1988), Flap and De Graaf (1989), and Marsden and Hurlbert (1988). Coleman (1988) discusses the transmission of human capital across generations. Flap and Tazelaar (1989) provide a thorough review with special attention to social network analysis.

Empirical questions in this line of work concern the magnitude of association between contact resources and the actor's own resources, and variation in the association across kinds of relationships. Granovetter's (1973) weak tie metaphor, discussed in detail shortly, is often invoked to distinguish kinds of relationships.²

Network analysts will recognize this as an example of social contagion analysis. Network structure is not used to predict attitudes or behaviors directly. It is used to predict similarity between attitudes and behaviors (compare Barber, 1978, for a causal analysis). The research tradition is tied to the Columbia Sociology survey studies of social influence conducted during the 1940s and 1950s. In one of the first well-known studies, for example, Lazarsfeld, Berelson, and Gaudet (1944) show how a person's vote is associated with the party affiliations of friends. Persons claiming to have voted for the presidential candidate of a specific political

party tend to have friends affiliated with that party. Social capital theory developed from this line of work describes the manner in which resources available to any one person in a population are contingent on the resources available to individuals socially proximate to the person.

Empirical evidence is readily available. People develop relations with people like themselves (for example, Fischer, 1982; Marsden, 1987; Burt, 1990b). Wealthy people develop ties with other wealthy people. Educated people develop ties with one another. Young people develop ties with one another. There are reasons for this. Socially similar people, even in the pursuit of independent interests, spend time in the same places. Relationships emerge. Socially similar people have more shared interests. Relationships are maintained. Further, we are sufficiently egocentric to find people with similar tastes attractive. Whatever the etiology for strong relations between socially similar people, it is to be expected that the resources and opinions of any one individual will be correlated with the resources and opinions of his or her close contacts.

How

A second line of work describes social structure as capital in its own right. The first line describes the network as a conduit; the second line describes how networks are themselves a form of social capital. This line of work is less developed than the first. Indeed, it is little developed beyond intuitions in empirical research on social capital. Network range, indicated by size, is the primary measure. For example, Boxman, De Graaf, and Flap (1991) show that people with larger contact networks obtain higher-paying positions than people with small networks. A similar finding in social support research shows that persons with larger networks tend to live longer (Berkman and Syme, 1979).

Both lines of work are essential to a general definition of social capital. Social capital is at once the resources contacts hold and the structure of contacts in a network. The first term describes whom you reach. The second describes how you reach.

For two reasons, however, I ignore the question of who to concentrate on how. The first is generality. The question of who elicits a more idiosyncratic class of answers. Predicting rate of return depends on knowing the resources of a player's contacts. There will be interesting empirical variation from one kind of activity to another, say, job searches versus mobilizing support for a charity, but the empirical generalization is obvious. Doing business with wealthy clients, however wealth is defined, has a higher margin than doing business with poor clients. I want to identify

parameters of social capital that generalize beyond the specific individuals connected by a relationship.

The second reason is correlation. The two components in social capital should be so strongly correlated that I can reconstruct much of the phenomenon from whichever component more easily yields a general explanation. To the extent that people play an active role in shaping their relationships, then a player who knows how to structure a network to provide high opportunity knows whom to include in the network. Even if networks are passively inherited, the manner in which a player is connected within social structure says much about contact resources. I will show that players with well-structured networks obtain higher rates of return. Resources accumulate in their hands. People develop relations with people like themselves. Therefore, how a player is connected in social structure indicates the volume of resources held by the player and the volume to which the player is connected.³

The nub of the matter is to describe network benefits in the competitive arena in order to be able to describe how certain structures enhance those benefits. The benefits are of two kinds, information and control.

Information

Opportunities spring up everywhere: new institutions and projects that need leadership, new funding initiatives looking for proposals, new jobs for which you know of a good candidate, valuable items entering the market for which you know interested buyers. The information benefits of a network define who knows about these opportunities, when they know, and who gets to participate in them. Players with a network optimally structured to provide these benefits enjoy higher rates of return to their investments, because such players know about, and have a hand in, more rewarding opportunities.

ACCESS, TIMING, AND REFERRALS

Information benefits occur in three forms: access, timing, and referrals. Access refers to receiving a valuable piece of information and knowing who can use it. Information does not spread evenly across the competitive arena. It isn't that players are secretive, although that too can be an issue. The issue is that players are unevenly connected with one another, are attentive to the information pertinent to themselves and their friends, and are all overwhelmed by the flow of information. There are limits to the volume of information you can use intelligently. You can only keep

up with so many books, articles, memos, and news services. Given a limit to the volume of information that anyone can process, the network becomes an important screening device. It is an army of people processing information who can call your attention to key bits—keeping you up to date on developing opportunities, warning you of impending disasters. This second-hand information is often fuzzy or inaccurate, but it serves to signal something to be looked into more carefully.

Related to knowing about an opportunity is knowing whom to bring into it. Given a limit to the financing and skills that we possess individually, most complex projects will require coordination with other people as staff, colleagues, or clients. The manager asks, “Whom do I know with the skills to do a good job with that part of the project?” The capitalist asks, “Whom do I know who would be interested in acquiring this product or a piece of the project?” The department head asks, “Who are the key players needed to strengthen the department’s position?” Add to each of these the more common question, “Whom do I know who is most likely to know the kind of person I need?”

Timing is a significant feature of the information received by network. Beyond making sure that you are informed, personal contacts can make you one of the people who is informed early. It is one thing to find out that the stock market is crashing today. It is another to discover that the price of your stocks will plummet tomorrow. It is one thing to learn the names of the two people referred to the board for the new vice-presidency. It is another to discover that the job will be created and that your credentials could make you a serious candidate for the position. Personal contacts get significant information to you before the average person receives it. That early warning is an opportunity to act on the information yourself or to invest it back into the network by passing it on to a friend who could benefit from it.

These benefits involve information flowing from contacts. There are also benefits in the opposite flow. The network that filters information coming to you also directs, concentrates, and legitimates information about you going to others.

In part, this network does no more than alleviate a logistics problem. You can only be in a limited number of places within a limited amount of time. Personal contacts get your name mentioned at the right time in the right place so that opportunities are presented to you. Their referrals are a positive force for future opportunities. They are the motor expanding the third category of people in your network, the players you don’t know who are aware of you. Consider the remark so often heard

in recruitment deliberations: “I don’t know her personally, but several people whose opinion I trust have spoken well of her.”

Beyond logistics, there is the issue of legitimacy. Even if you know about an opportunity and can present a solid case for why you should get it, you are a suspect source of information. The same information has more legitimacy when it comes from someone inside the decision-making process who can speak to your virtues. Candidates offered the university positions with the greatest opportunity, for example, are people who have a strong personal advocate in the decision-making process, a person in touch with the candidate to ensure that both favorable information and responses to any negative information get distributed during the decision.

BENEFIT-RICH NETWORKS

A player with a network rich in information benefits has contacts: (a) established in the places where useful bits of information are likely to air, and (b) providing a reliable flow of information to and from those places.

Selecting Contacts

The second criterion is as ambiguous as it is critical. It is a matter of trust, of confidence in the information passed and the care with which contacts look out for your interests. Trust is critical precisely because competition is imperfect. The question is not whether to trust, but whom to trust. In a perfectly competitive arena, you can trust the system to provide a fair return on your investments. In the imperfectly competitive arena, you have only your personal contacts. The matter comes down to a question of interpersonal debt. If I do for her, will she for me? There is no general answer. The answer lies in the match between specific people. If a contact feels that he is somehow better than you—a sexist male dealing with a woman, a racist white dealing with a black, an old-money matron dealing with an upwardly mobile ethnic—your investment in the relationship will be taken as proper obeisance to a superior. No debt is incurred. We use whatever cues can be found for a continuing evaluation of the trust in a relation, but we never know a debt is recognized until the trusted person helps us when we need it. With this kind of uncertainty, players are cautious about extending themselves for people whose reputation for honoring interpersonal debt is unknown. The importance of this point is illustrated by the political boundary around senior management discussed in Chapter 4. The more general point of trust as people meeting your expectations is illustrated in Barber’s (1983) analysis

of competence and duty as dimensions of trust relations in diverse institutions in American society.

Theory and research exist to identify trustworthy contacts. Strong relationships and mutual acquaintances tend to develop between people with similar social attributes such as education, income, occupation, and age (for example, Fischer, 1982; Burt, 1986, 1990b; Marsden, 1987; and see note 4 below). Both factors are linked to trust. Trust is a component in the strong relationships, and mutual acquaintances are like an insurance policy through which interpersonal debt is enforced such that the other person can be deemed trustworthy. (Nohria, 1991). Whether egocentrism, cues from presumed shared background and interests, or confidence in mutual acquaintances to enforce interpersonal debt, the operational guide to the formation of close, trusting relations seems to be that a person more like me is less likely to betray me. For the purposes here, I set the whole issue to one side as person-specific and presume that it is resolved by the able player.

Siting Contacts

That leaves the first criterion, establishing contacts where useful bits of information are likely to air. Everything else constant, a large, diverse network is the best guarantee of having a contact present where useful information is aired. This is not to say that benefits must increase linearly with size and diversity, a point to which I will return (Figure 1.5), but only that, other things held constant, the information benefits of a large, diverse network are more than the information benefits of a small, homogeneous network.

Size is the more familiar criterion. Bigger is better. Acting on this understanding, people can expand their networks by adding more and more contacts. They make more cold calls, affiliate with more clubs, attend more social functions. Numerous books and self-help groups can assist them in “networking” their way to success by putting them in contact with a large number of potentially useful, or helpful, or like-minded people. The process is illustrated by the networks in Figure 1.1. The four-contact network at the left expands to sixteen contacts at the right. Relations are developed with a friend of each contact in network A, doubling the contacts to eight in network B. Snowballing through friends of friends, there are sixteen contacts in network C, and so on.

Size is a mixed blessing. More contacts can mean more exposure to valuable information, more likely early exposure, and more referrals.

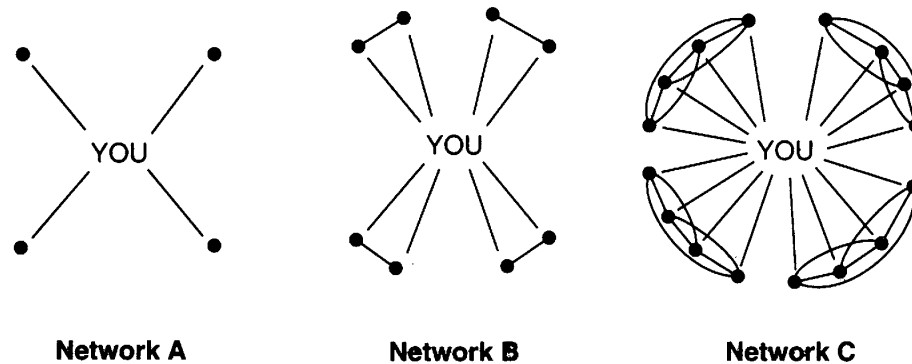


Figure 1.1 Network expansion

But increasing network size without considering diversity can cripple a network in significant ways. What matters is the number of nonredundant contacts. Contacts are redundant to the extent that they lead to the same people, and so provide the same information benefits.

Consider two four-contact networks, one sparse, the other dense. There are no relations between the contacts in the sparse network, and strong relations between every contact in the dense network. Both networks cost whatever time and energy is required to maintain four relationships. The sparse network provides four nonredundant contacts, one for each relationship. No single one of the contacts gets the player to the same people reached by the other contacts. In the dense network, each relationship puts the player in contact with the same people reached through the other relationships. The dense network contains only one nonredundant contact. Any three are redundant with the fourth.

The sparse network provides more information benefits. It reaches information in four separate areas of social activity. The dense network is a virtually worthless monitoring device. Because the relations between people in that network are strong, each person knows what the other people know and all will discover the same opportunities at the same time.

The issue is opportunity costs. At minimum, the dense network is inefficient in the sense that it returns less diverse information for the same cost as that of the sparse network. A solution is to put more time and energy into adding nonredundant contacts to the dense network. But time and energy are limited, which means that inefficiency translates into opportunity costs. If I take four relationships as an illustrative limit on the number of strong relations that a player can maintain, the player in the dense network is cut off from three fourths of the information provided by the sparse network.

Structural Holes

I use the term structural hole for the separation between nonredundant contacts. Nonredundant contacts are connected by a structural hole. A structural hole is a relationship of nonredundancy between two contacts. The hole is a buffer, like an insulator in an electric circuit. As a result of the hole between them, the two contacts provide network benefits that are in some degree additive rather than overlapping.

EMPIRICAL INDICATORS

Nonredundant contacts are disconnected in some way—either directly, in the sense that they have no direct contact with one another, or indirectly, in the sense that one has contacts that exclude the others. The respective empirical conditions that indicate a structural hole are cohesion and structural equivalence. Both conditions define holes by indicating where they are absent.

Under the cohesion criterion, two contacts are redundant to the extent that they are connected by a strong relationship. A strong relationship indicates the absence of a structural hole. Examples are father and son, brother and sister, husband and wife, close friends, people who have been partners for a long time, people who frequently get together for social occasions, and so on. You have easy access to both people if either is a contact. Redundancy by cohesion is illustrated at the top of Figure 1.2. The three contacts are connected to one another, and so provide

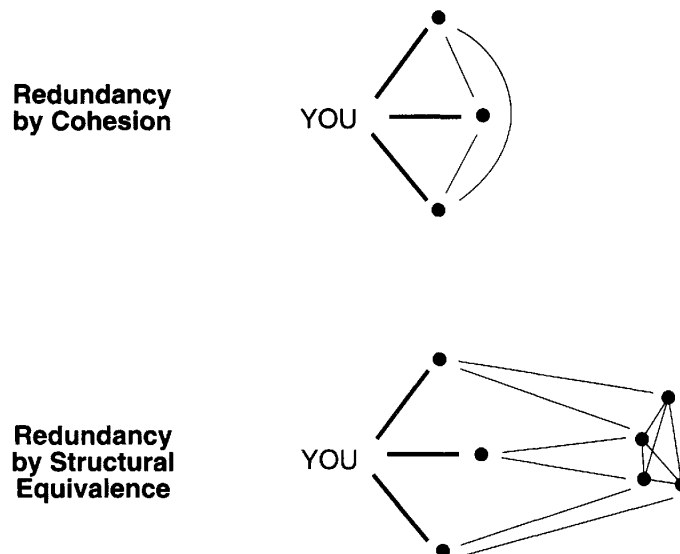


Figure 1.2 Structural indicators of redundancy

the same network benefits. The presumption here—routine in network analysis since Festinger, Schachter, and Back's (1950) analysis of information flowing through personal relations and Homans's (1950) theory of social groups—is that the likelihood that information will move from one person to another is proportional to the strength of their relationship. Empirically, strength has two independent dimensions: frequent contact and emotional closeness (see Marsden and Hurlbert, 1988; Burt, 1990b).

Structural equivalence is a useful second indicator for detecting structural holes. Two people are structurally equivalent to the extent that they have the same contacts. Regardless of the relation between structurally equivalent people, they lead to the same sources of information and so are redundant. Cohesion concerns direct connection; structural equivalence concerns indirect connection by mutual contact. Redundancy by structural equivalence is illustrated at the bottom of Figure 1.2. The three contacts have no direct ties with one another. They are nonredundant by cohesion. But each leads you to the same cluster of more distant players. The information that comes to them, and the people to whom they send information, are redundant. Both networks in Figure 1.2 provide one nonredundant contact at a cost of maintaining three.

The indicators are neither absolute nor independent. Relations deemed strong are only strong relative to others. They are our strongest relations. Structural equivalence rarely reaches the extreme of complete equivalence. People are more or less structurally equivalent. In addition, the criteria are correlated. People who spend a lot of time with the same other people often get to know one another. The mutual contacts responsible for structural equivalence set a stage for the direct connection of cohesion. The empirical conditions between two players will be a messy combination of cohesion and structural equivalence, present to varying degrees, at varying levels of correlation.

Cohesion is the more certain indicator. If two people are connected with the same people in a player's network (making them redundant by structural equivalence), they can still be connected with different people beyond the network (making them nonredundant). But if they meet frequently and feel close to one another, then they are likely to communicate and probably have contacts in common. More generally, and especially for field work informed by attention to network benefits, the general guide is the definition of a structural hole. There is a structural hole between two people who provide nonredundant network benefits. If the cohesion and structural equivalence conditions are considered together, redundancy is most likely between structurally equivalent people connected by

a strong relationship. Redundancy is unlikely, indicating a structural hole, between total strangers in distant groups. I will return to this issue again, to discuss the depth of a hole, after control benefits have been introduced.

THE EFFICIENT-EFFECTIVE NETWORK

Balancing network size and diversity is a question of optimizing structural holes. The number of structural holes can be expected to increase with network size, but the holes are the key to information benefits. The optimized network has two design principles.

Efficiency

The first design principle of an optimized network concerns efficiency: Maximize the number of nonredundant contacts in the network to maximize the yield in structural holes per contact. Given two networks of equal size, the one with more nonredundant contacts provides more benefits. There is little gain from a new contact redundant with existing contacts. Time and energy would be better spent cultivating a new contact to unreached people.⁴ Maximizing the nonredundancy of contacts maximizes the structural holes obtained per contact.⁵

Efficiency is illustrated by the networks in Figure 1.3. These reach the same people reached by the networks in Figure 1.1, but in a different way. What expands in Figure 1.1 is not the benefits, but the cost of maintaining the network. Network A provides four nonredundant contacts. Network B provides the same number. The information benefits provided by the initial four contacts are redundant with benefits provided by their close friends. All that has changed is the doubled number of relationships maintained in the network. The situation deteriorates even further with the sixteen contacts in network C. There are still only four

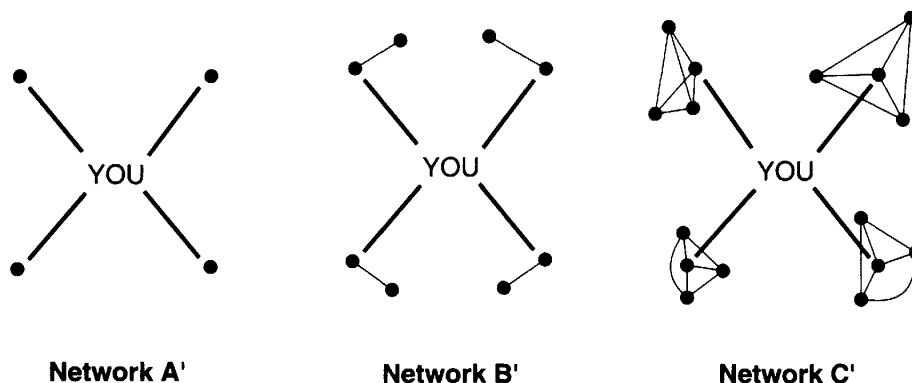


Figure 1.3 Strategic network expansion

nonredundant contacts in the network, but their benefits are now obtained at a cost of maintaining sixteen relationships.

With a little network surgery, the sixteen contacts can be maintained at a fourth of the cost. As illustrated in Figure 1.3, select one contact in each cluster to be a primary link to the cluster. Concentrate on maintaining the primary contact, and allow direct relationships with others in the cluster to weaken into indirect relations through the primary contact. These players reached indirectly are secondary contacts. Among the redundant contacts in a cluster, the primary contact should be the one most easily maintained and most likely to honor an interpersonal debt to you in particular. The secondary contacts are less easily maintained or less likely to work for you (even if they might work well for someone else). The critical decision obviously lies in selecting the right person to be a primary contact. The importance of trust has already been discussed. With a trustworthy primary contact, there is little loss in information benefits from the cluster and a gain in the reduced effort needed to maintain the cluster in the network.

Repeating this operation for each cluster in the network recovers effort that would otherwise be spent maintaining redundant contacts. By reinvesting that saved time and effort in developing primary contacts to new clusters, the network expands to include an exponentially larger number of contacts while expanding contact diversity. The sixteen contacts in network C of Figure 1.1, for example, are maintained at a cost of four primary contacts in network C' of Figure 1.3. Some portion of the time spent maintaining the redundant other twelve contacts can be reallocated to expanding the network to include new clusters.

Effectiveness

The second design principle of an optimized network requires a further shift in perspective: Distinguish primary from secondary contacts in order to focus resources on preserving the primary contacts. Here contacts are not people on the other end of your relations; they are ports of access to clusters of people beyond. Guided by the first principle, these ports should be nonredundant so as to reach separate, and therefore more diverse, social worlds of network benefits. Instead of maintaining relations with all contacts, the task of maintaining the total network is delegated to primary contacts. The player at the center of the network is then free to focus on properly supporting relations with primary contacts and expanding the network to include new clusters. The first principle concerns the average number of people reached with a primary contact; the

second concerns the total number of people reached with all primary contacts. The first principle concerns the yield per primary contact. The second concerns the total yield of the network. More concretely, the first principle moves from the networks in Figure 1.1 to the corresponding networks in Figure 1.3. The second principle moves from left to right in Figure 1.3. The target is network C' in Figure 1.3: a network of few primary contacts, each a port of access to a cluster of many secondary contacts.

Figure 1.4 illustrates some complexities in unpacking a network to maximize structural holes. The “before” network contains five primary contacts and reaches a total of fifteen people. However, there are only two clusters of nonredundant contacts in the network. Contacts 2 and 3 are redundant in the sense of being connected with each other and reaching the same people (cohesion and structural equivalence criteria). The same is true of contacts 4 and 5. Contact 1 is not connected directly to contact 2, but he reaches the same secondary contacts; thus contacts 1 and 2 provide redundant network benefits (structural equivalence crite-

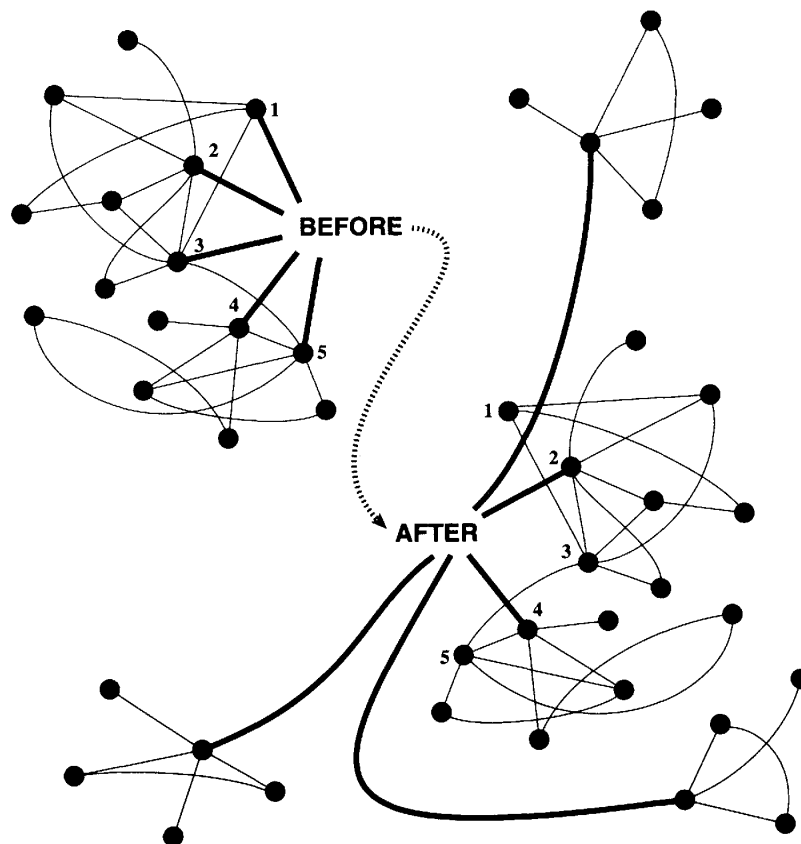


Figure 1.4 Optimizing for structural holes

tion). Illustrating the other extreme, contacts 3 and 5 are connected directly, but they are nonredundant because they reach separate clusters of secondary contacts (structural equivalence criterion). In the “after” network, contact 2 is used to reach the first cluster in the “before” network and contact 4 is used to reach the second cluster. The time and energy saved by withdrawing from relations with the other three primary contacts is reallocated to primary contacts in new clusters. The “before” and “after” networks are both maintained at a cost of five primary relationships, but the “after” network is dramatically richer in structural holes, and so network benefits.

Network benefits are enhanced in several ways. There is a higher volume of benefits, because more contacts are included in the network. Beyond volume, diversity enhances the quality of benefits. Nonredundant contacts ensure exposure to diverse sources of information. Each cluster of contacts is an independent source of information. One cluster, no matter how numerous its members, is only one source of information, because people connected to one another tend to know about the same things at about the same time. The information screen provided by multiple clusters of contacts is broader, providing better assurance that you, the player, will be informed of opportunities and impending disasters. Further, because nonredundant contacts are only linked through the central player, you are assured of being the first to see new opportunities created by needs in one group that could be served by skills in another group. You become the person who first brings people together, which gives you the opportunity to coordinate their activities. These benefits are compounded by the fact that having a network that yields such benefits makes you even more attractive as a network contact to other people, thus easing your task of expanding the network to best serve your interests.

Growth Patterns

A more general sense of efficiency and effectiveness is illustrated with network growth. In Figure 1.5, the number of contacts in a player’s network increases from left to right on the horizontal axis. The number who are nonredundant increases up the vertical axis. Observed network size increases on the horizontal, effective size up the vertical. Networks can be anywhere in the gray area. The maximum efficiency line describes networks in which each new contact is completely nonredundant with other contacts. Effective size equals actual size. Efficient-effective networks are in the upper right of the graph. The minimum efficiency line

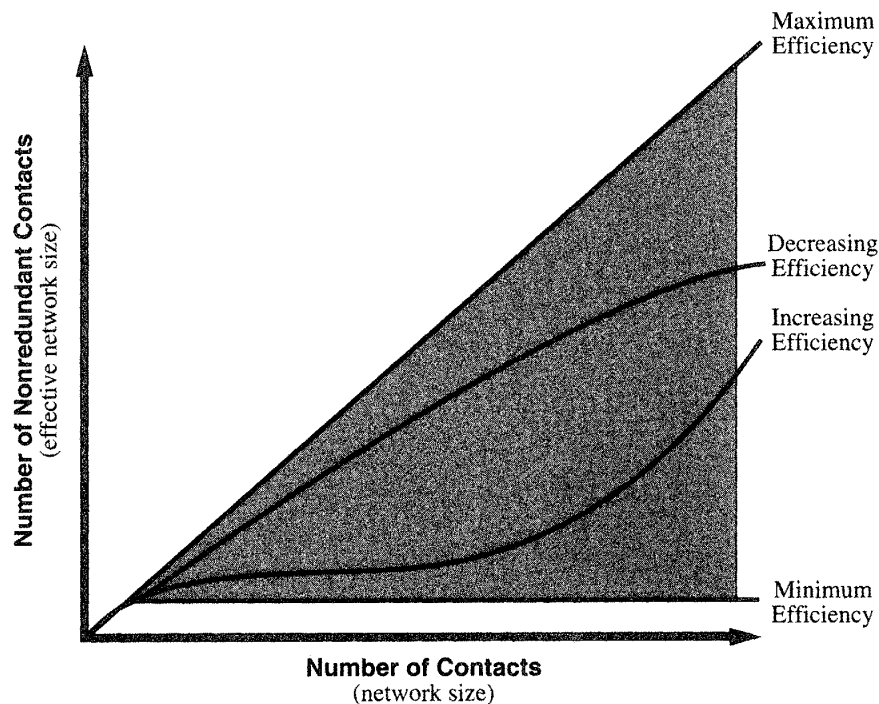


Figure 1.5 Efficiency and effectiveness

describes networks in which each new contact is completely redundant with other contacts; effective size equals one, regardless of multiple contacts in the network.

The two lines between the extremes illustrate more probable growth patterns. The decreasing efficiency line shows players building good information benefits into their initial network, then relaxing to allow increasing redundancy as the network gets large. Friends of friends begin to be included. Comparisons across networks of different sizes suggest that this is the growth pattern among managers (see Figure 4.15), though controls for time would be necessary to make the suggestion an inference.

The increasing efficiency line illustrates a different growth pattern. Initial contacts are redundant with one another. A foundation is established with multiple contacts in the same cluster. After the foundation is established, the player's network expands to include contacts in other clusters and effective size begins to increase. There are two kinds of clusters in which optimizing for saturation is wiser than optimizing for efficiency. The first is obvious. Leisure and domestic clusters are a congenial environment of low-maintenance, redundant contacts. Efficiency mixes poorly with friendship. Judging friends on the basis of efficiency is an

interpersonal flatulence from which friends will flee. The second exception is a cluster of contacts where resources are dense. For the CEO, the board of directors is such a cluster. The university provost is similarly tied to the board of trustees. For the more typical manager, the immediate work group is such a cluster, especially with respect to funding authority within the group. These clusters are so important to the vitality of the rest of the network that it is worth treating each person in them as a primary contact, regardless of redundancy. Saturation minimizes the risk of losing effective contact with the cluster and minimizes the risk of missing an important opportunity anywhere in the cluster.

The more general point is that the probability of receiving network benefits from a cluster has two components, the probability that a contact will transmit information to you and the probability that it will be transmitted to the contact. I count on dense ties within a cluster to set the second probability to one. The probability of having a benefit transmitted to you therefore depends only on the strength of your relationship with a contact in the cluster. However, where the density of ties in an opportunity-rich cluster lowers the probability that your contact will know about an opportunity, there is value in increasing the number, and thus the redundancy, of contacts in the cluster so that total coverage of the cluster compensates for imperfect transmission within it.

STRUCTURAL HOLES AND WEAK TIES

Discontinuities in social structure have long been a subject of study in sociology. Fitting the structural hole argument into the history of sociological thought is not the task of this book, but one piece of contemporary history adds value to the argument here. Mark Granovetter's weak tie argument provides an illuminating aside on the information benefits of structural holes.

History

In the late 1960s and early 1970s at Harvard University, Harrison White, with a cluster of exceptional sociology graduate students, was engaged in studying the importance of gaps, as opposed to the ties, in social structure. First came his celebrated work on chains of mobility (White, 1970), and later his work with colleagues, most notably Ronald Breiger and Scott Boorman, on concrete network models—blockmodels—of social structure (White, Boorman, and Breiger, 1976; see Burt, 1982:63–69, for review). The usual analysis of mobility describes patterns of mobility, or careers, created by people moving between positions in a social struc-

ture. White (1970) shifted perspective to focus on the hole, or opportunity, created when a person leaves a position. As people move up the hierarchy, they create opportunities for people below them. Chains of promotion move up a hierarchy. Chains of opportunity move down. Looking at social structure more generally, White, Boorman, and Breiger (1976, esp. pp. 732n, 737–740) stressed the structural hole metaphor as a substantive motivation for their network blockmodels. They focused on “zeroblocks” as an especially significant component in the relation pattern defining a position in social structure. It is clear from their analysis that they meant structural holes to be important for understanding network contingent action as well as the task they addressed of clustering network elements into blocks (for example, see pp. 763ff. on the low rate of change in zeroblocks).

One of the students, Mark Granovetter, found a troubling result in his dissertation research. Hoping to link network structure to job searches, he interviewed men about how they found their current jobs and included sociometric items asking for the names of close contacts. The troubling result was that the men almost never found work through close contacts. When information on a job opportunity came through a personal contact, the contact was often distant, such as a high school acquaintance met by accident at a recent social event. He developed the point in a widely cited article, “The Strength of Weak Ties” (Granovetter, 1973), and in a book, *Getting a Job* (Granovetter, 1974).

Connecting the Two Arguments

The weak tie argument is elegantly simple. The stage is set with results familiar from the social psychology of Festinger and Homans circa 1950, discussed above with respect to cohesion indicators of structural holes. People live in a cluster of others with whom they have strong relations. Information circulates at a high velocity within these clusters. Each person tends to know what the other people know. The spread of information on new ideas and opportunities, therefore, must come through the weak ties that connect people in separate clusters. The weak ties so often ignored by social scientists are in fact a critical element of social structure. Hence the strength of weak ties. Weak ties are essential to the flow of information that integrates otherwise disconnected social clusters into a broader society.

The idea and its connection with structural holes is illustrated in Figure 1.6. There are three clusters of players. Strong ties, indicated by solid lines, connect players within clusters. Dashed lines indicate two weak

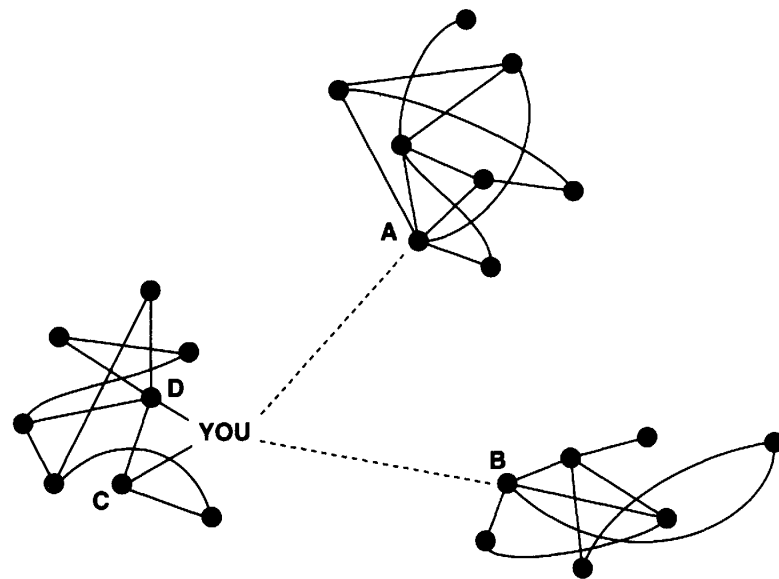


Figure 1.6 Structural holes and weak ties

ties between players in separate clusters. One of the players, you, has a unique pattern of four ties: two strong ties within your cluster and a weak tie to a contact in each in the other clusters. There are three classes of structural holes in your network: (a) holes between the cluster around contact A and everyone in your own cluster, for example, the hole between contacts A and C; (b) holes between the cluster around contact B and everyone in your own cluster, for example, the hole between contacts B and C; and (c) the hole between contacts A and B.

Weak ties and structural holes seem to describe the same phenomenon. In Figure 1.6, for example, they predict the same ranking of information benefits. You are best positioned for information benefits, contacts A and B are next, followed by everyone else. You have two weak ties, contacts A and B have one each, and everyone else has none. You have the largest volume of structural holes between your contacts, contacts A and B have fewer, and everyone else has few or none.

The Strength of Structural Holes

The weak tie argument is simpler and already well known. Why complicate the situation with the structural hole argument? There are two reasons.

First, the causal agent in the phenomenon is not the weakness of a tie but the structural hole it spans. Tie weakness is a correlate, not a cause. The structural hole argument captures the causal agent directly and thus

provides a stronger foundation for theory and a clearer guide for empirical research. Second, by shifting attention away from the structural hole responsible for information benefits, to the strength of the tie providing them, the weak tie argument obscures the control benefits of structural holes. Control benefits augment and in some ways are more important than the information benefits of structural holes. Building both benefits into the argument speaks more clearly to the generality of the phenomenon under study. I will elaborate the first point, then move to the second in the next section.

The weak tie argument is about the strength of relationships at the same time that it is about their location. The two dashed lines in Figure 1.6 are bridges. They are the only connection between two otherwise separate clusters of strongly interconnected players (compare Granovetter, 1973:1065, on weak ties as bridges). A bridge is at once two things. It is a chasm spanned and the span itself. By title and subsequent application, the weak tie argument is about the strength of relationships that span the chasm between two social clusters. The structural hole argument is about the chasm spanned. It is the latter that generates information benefits. Whether a relationship is strong or weak, it generates information benefits when it is a bridge over a structural hole.

Consider a crosstabulation of ties by their strength and location. Your relationships can be sorted into two categories of strength. Strong ties are your most frequent and close contacts. Weak ties are your less frequent, less close contacts. Between these two categories, you have a few strong ties and many weak ties.

Now sort, by location, redundant ties within your social cluster versus nonredundant ties to people in other clusters. The nonredundant ties are your bridges to other clusters. From what we know about the natural etiology of relationships, bridges are less likely to develop than ties within clusters. The category of redundant ties includes your strong ties to close friends and colleagues, whom you see often, but it also includes their friends, and friends of friends, whom you meet only occasionally if at all. As you expand your inventory from your closest, most frequent contacts to your more distant, contacts tend to be people like yourself before you reach a sufficiently low level of relationship to include people from completely separate social worlds. This tendency varies from one person to the next, but it is in the aggregate the substance of the well-documented tendency already discussed for relations to develop between socially similar people. In Figure 1.6, you are one of nine people in your social cluster. You have strong ties to two people. Through those two, you have weak

ties to the other six people in the cluster. To keep the sociogram simple, I deleted the dashed lines for those ties and their equivalent inside the other clusters. The other six people in your cluster are friends of friends whom you know and sometimes meet but don't have the time or energy to include among your closest contacts. The cluster is clearly held together by strong ties. Everyone has two to five strong ties to others within the cluster. All nine people are likely to know about the same opportunities as expected in a cohesive cluster. Of the 36 possible connections among the nine people in the cluster, however, only 12 are solid line strong ties. The remaining two thirds are weak ties between redundant friends of friends.

Now crosstabulate the two classifications and take expected values. The result is given in Table 1.1. Information benefits vary across the columns of the table and are higher through nonredundant ties. This is accurately represented in both the weak tie and the structural hole argument. But a quick reading of the weak tie argument, with its emphasis on the strength of a relationship, has led some to test the idea that information benefits covary inversely with the strength of ties. This is a correlation between the rows and columns of Table 1.1, which is no correlation at all. In fact, the typical tie in Table 1.1 is weak and provides redundant information. The correlation in a study population depends on the distribution of ties in the table, but there is no theoretical reason to expect a strong correlation between the strength of a relationship and the information benefits it provides.

The weak tie argument is about the two cells in the second column of the table. It predicts that nonredundant ties, the bridges that provide information benefits, are more likely weak than strong. In the second column of Table 1.1, weak tie bridges are more likely than strong tie bridges. To simplify his argument, Granovetter makes this tendency absolute by ruling out strong tie bridges (the "rare" cell in Table 1.1, the

Table 1.1 The natural distribution of relationships

| Strength | Location in social structure | | TOTAL |
|------------|------------------------------|---------------------------------|-------|
| | Redundant tie within cluster | Nonredundant tie beyond cluster | |
| Weak tie | many | some | MORE |
| Strong tie | some | rare | LESS |
| TOTAL | MORE | LESS | |

“forbidden triad” in Granovetter’s argument, 1973:1063). He (1973:1064) says, “A strong tie can be a bridge, therefore, only if neither party to it has any other strong ties, unlikely in a social network of any size (though possible in a small group). Weak ties suffer no such restriction, though they are certainly not automatically bridges. What is important, rather, is that all bridges are weak ties.”

Bridge strength is an aside in the structural hole argument. Information benefits are expected to travel over all bridges, strong or weak. Benefits vary between redundant and nonredundant ties, the columns of Table 1.1. Thus structural holes capture the condition directly responsible for the information benefits. The task for a strategic player building an efficient-effective network is to focus resources on the maintenance of bridge ties. Otherwise, and this is the correlative substance of the weak tie argument, bridges will fall into their natural state of being weak ties.

Control and the Tertius Gaudens

I have described how structural holes can determine who knows about opportunities, when they know, and who gets to participate in them. Players with a network optimized for structural holes, in addition to being exposed to more rewarding opportunities, are also more likely to secure favorable terms in the opportunities they choose to pursue. The structural holes that generate information benefits also generate control benefits, giving certain players an advantage in negotiating their relationships. To describe how this is so, I break the negotiation into structural, motivational, and outcome components (corresponding to the textbook distinction between market structure, market conduct, and market performance; for example, Caves, 1982). The social structure of the competitive arena defines opportunities, a player decides to pursue an opportunity, and is sometimes successful. I will begin with the outcome.

TERTIUS GAUDENS

Sometimes you will emerge successful from negotiation as the *tertius gaudens*. Taken from the work of Georg Simmel, the *tertius* role is useful here because it defines successful negotiation in terms of the social structure of the situation in which negotiation is successful. The role is the heart of Simmel’s (1922) later analysis of the freedom an individual derives from conflicting group affiliations (see Coser, 1975, for elaboration).⁶ The *tertius gaudens* is “the third who benefits” (Simmel, 1923:154, 232).⁷ The phrase survives in an Italian proverb, *Far i due litiganti, il terzo*

gode (Between two fighters, the third benefits), and, to the north, in a more jovial Dutch wording, *de lachende derde* (the laughing third).⁸ *Tertius*, *terzo*, or *derde*, the phrase describes an individual who profits from the disunion of others.

There are two *tertius* strategies: being the third between two or more players after the same relationship, and being the third between players in two or more relations with conflicting demands. The first, and simpler, strategy is the familiar one that occurs in economic bargaining between buyer and seller. When two or more players want to buy something, the seller can play their bids against one another to get a higher price. The strategy extends directly: a woman with multiple suitors or a professor with simultaneous offers of positions in rival institutions.

The control benefits of having a choice between players after the same relationship extends directly to choice between the simultaneous demands of players in separate relationships. The strategy can be seen between hierarchical statuses in the enterprising subordinate under the authority of two or more superiors: for example, the student who strikes her own balance between the simultaneous demands of imperious faculty advisors.⁹ The bargaining is not limited to situations of explicit competition. In some situations, emerging as the *tertius* depends on creating competition. In proposing the concept of a role-set, for example, Merton (1957:393–394) identifies this as a strategy to resolve conflicting role demands. Make simultaneous, contradictory demands explicit to the people posing them, and ask them to resolve their—now explicit—conflict. Even where it doesn't exist, competition can be produced by defining issues such that contact demands become contradictory and must be resolved before you can meet their requests. Failure is possible. You might provide too little incentive for the contacts to resolve their differences. Contacts drawn from different social strata need not perceive one another's demands as carrying equal weight. Or you might provide too much incentive. Now aware of one another, the contacts could discover sufficient reason to cooperate in forcing you to meet their mutually agreed-upon demands (Simmel, 1902:176, 180–181, calls attention to such failures). But if the strategy is successful, the pressure on you is alleviated and is replaced with an element of control over the negotiation. Merton (1957:430) states the situation succinctly: the player at the center of the network, “originally at the focus of the conflict, virtually becomes a more or less influential bystander whose function it is to high-light the conflicting demands by members of his role-set and to make it a problem for them, rather than for him, to resolve *their* contradictory demands.”

The strategy holds equally well with large groups. Under the rubric

“divide and rule,” Simmel (1902:185–186) describes institutional mechanisms through which the Incan and Venetian governments obtained advantage by creating conflict between subjects. The same point is illustrated more richly in Barkey’s (1991) comparative description of state control in early seventeenth-century France and Turkey. After establishing the similar conditions in the two states at the time, Barkey asks why peasant-noble alliances developed in France against the central state while no analogous or substitutable alliances developed in Turkey. The two empires were comparable with respect to many factors that scholars have cited to account for peasant revolt. They differed in one significant factor correlated with revolt—not in the structure of centralized state control, but in control strategy. In France, the king sent trusted representatives as agents to collect taxes and to carry out military decisions in provincial populations. These outside agents, *intendants*, affected fundamental local decisions and their intrusion was resented by the established local nobility. Local nobility formed alliances with the peasantry against the central state. In Turkey, the sultan capitalized on conflict among leaders in the provinces. When a bandit became a serious threat to the recognized governor, a deal was struck with the bandit to make him the legitimate governor. Barkey (1991:710) writes: “At its extreme, the state could render a dangerous rebel legitimate overnight by striking a bargain that ensured new sources of revenue for the rebel and momentary relief from internal warfare and, perhaps, an army or two for the state.” The two empires differed in their use of structural holes. The French king, assuming he had absolute authority, ignored them. The Turkish sultan, promoting competition between alternative leaders, strategically exploited them. Conflict within the Turkish empire remained in the province, rather than being directed against the central state. As is characteristic of the control obtained via structural holes, the resulting Turkish control was more negotiated than was the absolute control exercised in France. It was also more effective.

THE ESSENTIAL TENSION

There is a presumption of tension here. Control emerges from *tertius* brokering tension between other players. No tension, no *tertius*.

It is easy to infer that the tension presumed is the tension between combatants. There is certainly a *tertius*-rich tension between combatants. Governors and bandits in the Turkish game played for life or death stakes. A corporate executive listening to the control argument illustrates the problem. Her colleagues, she explained, took pride in working together

in a spirit of partnership and goodwill. The *tertius* imagery rang true for many firms she knew of, but not her own.

The reasoning is good. The conclusion is wrong. I referred the skeptical executive to Chapter 4, which by coincidence is an analysis of managers at her level, in her firm. Promotions in the firm are strongly correlated, and illuminatingly so for women, with the structural holes in a manager's network.

The tension essential to the *tertius* is merely uncertainty. Separate the uncertainty of control from its consequences. The consequences can be life or death, in the extreme, or merely a question of embarrassment. Everyone knows you made an effort to get that job, but it went to someone else. The *tertius* strategies can be applied to control with severe consequences or to control of little consequence. What is essential is that control is uncertain, that no one can act as if he or she has absolute authority. Where there is any uncertainty about whose preferences should dominate a relationship, there is an opportunity for the *tertius* to broker the negotiation for control by playing demands against one another. There is no long-term contract that keeps a relationship strong, no legal binding that can secure the trust necessary to a productive relationship. Your network is a pulsing swirl of mixed, conflicting demands. Each contact wants your exclusive attention, your immediate response when a concern arises. All, to warrant their continued confidence in you, want to see you measure up to the values against which they judge themselves. Within this preference webwork, where no demands have absolute authority, the *tertius* negotiates for favorable terms.

THE CONNECTION WITH INFORMATION BENEFITS

Structural holes are the setting for *tertius* strategies. Information is the substance. Accurate, ambiguous, or distorted information is moved between contacts by the *tertius*. One bidder is informed of a competitive offer in the first *tertius* strategy. A player in one relationship is informed of demands from other relationships in the second *tertius* strategy.

The two kinds of benefits augment and depend on one another. Application of the *tertius* strategies elicits additional information from contacts interested in resolving the negotiation in favor of their own preferences. The information benefits of access, timing, and referrals enhance the application of strategy. Successful application of the *tertius* strategies involves bringing together players who are willing to negotiate, have sufficiently comparable resources to view one another's preferences as valid, but won't negotiate with one another directly to the exclusion of the *tertius*. Having access to information means being able to identify

where there will be an advantage in bringing contacts together and is the key to understanding the resources and preferences being played against one another. Having that information early is the difference between being the one who brings together contacts versus being just another person who hears about the negotiation. Referrals further enhance strategy. It is one thing to distribute information between two contacts during negotiation. It is another thing to have people close to each contact endorsing the legitimacy of the information you distribute.

Entrepreneurs

I have described how the information and control benefits that are relevant to gaining an advantage in negotiating relationships are multiplicative. They augment and depend on one another, and together emerge from the wellspring of structural holes in a network. But what prompts a player to pursue these benefits? Negotiation contains a motivational component.

THE ISSUE OF MOTIVATION

Behavior of a specific kind converts opportunity into higher rates of return. The information benefits of structural holes might come to a passive player, but control benefits require an active hand in the distribution of information. Motivation is now an issue. Knowing about an opportunity and being in a position to develop it are distinct from doing something about it. The *tertius* plays conflicting demands and preferences against one another and builds value from their disunion. You enter the structural hole between two players to broker the relationship between them. Such behavior is not to everyone's taste. A player can respond in ways ranging from fully developing the opportunity to ignoring it. When you take the opportunity to be the *tertius*, you are an entrepreneur in the literal sense of the word—a person who generates profit from being between others. Both terms will be useful in these precise meanings; entrepreneur refers to a kind of behavior, the *tertius* is a successful entrepreneur.¹⁰

Both are distinct from behavior subsequent to emerging as the *tertius*. The *tertius* can choose to extract value from negotiated relations, or to add value, strengthening the relations for later profit. Some reinvestment is to be expected if the player's network is to remain intact. A nonprofit player, pursuing entrepreneurial opportunities just for the pleasure of being the one who brings others together to build value, could choose to

reinvest it all. The issue at hand is not the uses to which profit is put. It is who chooses to have a hand in the distribution of profit.

Motivation can be traced to cultural images of good and evil. In *The Protestant Ethic and the Spirit of Capitalism*, Weber (1905, esp. pp. 166ff.) describes the seventeenth-century bourgeois Protestant as an individual seeking—in his religious duty, his Calvinist “calling”—the profit of sober, thrifty, diligent exploitation of opportunities for usury and trade. Kilby (1971) provides a review and criticism of research on culturally induced entrepreneurs.

Psychological need is another motive. McClelland (1961) describes the formation in childhood of a need to achieve as critical to later entrepreneurial behavior (a need that can also be cultivated later if desired, McClelland, 1975). Without going into the etiology of motive, Schumpeter (1912:93) stresses nonutilitarian motives for entrepreneurship: “First of all, there is the dream and the will to found a private kingdom, usually, though not necessarily, also a dynasty . . . Then there is the will to conquer: the impulse to fight, to prove oneself superior to others, to succeed for the sake, not of the fruits of success, but of the success itself . . . Finally, there is the joy of creating, of getting things done, or simply of exercising one’s energy and ingenuity.”¹¹

OPPORTUNITY AND MOTIVATION

These are powerful frameworks for understanding competition, but I don’t wish to detour into the beliefs behind entrepreneurial behavior. I propose to leap over the motivation issue by taking the network as simultaneously an indicator of entrepreneurial opportunity and of motivation. Psychological and cultural motives for entrepreneurial behavior have been conceptualized and studied without data on the social network surrounding the entrepreneur. Such data are the substance of the structural hole argument and, in three ways, carry their own answer to the question of motivation.

First, there is the clarity of an opportunity. The above are “push” explanations. Players are pushed by psychological need or cultural imperative to be entrepreneurs. There is also a “pull” explanation. Players can be pulled to entrepreneurial action by the promise of success. I do not mean that players are rational creatures expected to calculate accurately and act in their own interest. Nor do I mean to limit the scope of the argument to situations in which players act as if they are rational in that way. I mean simply that given two opportunities, any player is more likely to act on the one with the clearer path to success. The clarity

of opportunity is its own motivation. As the number of entrepreneurial opportunities in a network increases, the odds of some being clearly defined by deep structural holes increases, and therefore the odds of entrepreneurial behavior increase. To be sure, a person whose abilities or values proscribe entrepreneurial behavior is unlikely to act, and someone inclined to entrepreneurial behavior is more likely to act or even to take the initiative to create opportunities.¹² Regardless of ability or values, however, within the broad range of acceptable behaviors, a person is unlikely to take entrepreneurial action if the probability of success is low. An observer might question the propriety of a scholar who negotiates with several universities offering a position, but the question is not an issue for the player with one offer.

There are also network analogues to the push explanations of motive. A person with a psychological need for entrepreneurial behavior is prone to building a network configured around such behavior. If I find a player with a network rich in the structural holes that make entrepreneurial behavior possible, I have a player willing and able to act entrepreneurially. But it is the rare person who is the sole author of a network. Networks are more often built in the course of doing something else. If your work, for example, involves meeting people from different walks of life, your network will end up composed of contacts who without you have no contact with one another. Even so, the network is its own explanation of motive. As the volume of structural holes in a player's network increases—regardless of the process that created them—the entrepreneurial behavior of making and negotiating relations between others becomes a way of life. This is a network analogue to the cultural explanation of motive. If all you know is entrepreneurial relationships, the motivation question is a nonissue. Being willing and able to act entrepreneurially is how you understand social life.

I will treat motivation and opportunity as one and the same. For reasons of a clear path to success, or the tastes of the player as the network's author, or the nature of the player's environment as author of the network, a network rich in entrepreneurial opportunity surrounds a player motivated to be entrepreneurial. At the other extreme, a player innocent of entrepreneurial motive lives in a network devoid of entrepreneurial opportunity.¹³

MEASUREMENT IMPLICATIONS

This detour into the issue of entrepreneurial motivation highlights a complexity that might otherwise obscure the association between structural

holes and rates of return. Consider Figure 1.7. Players are defined by their rate of return on investments (vertical axis) and the entrepreneurial opportunities of structural holes in their networks (horizontal axis).

The sloping line in the graph describes the hole effect of players rich in structural holes (horizontal axis) getting higher rates of return on investments (vertical axis). The increasingly positive slope of the line captures the increasing likelihood of *tertius* profit. A player invests in certain relationships. They need not all be high-yield relationships. The higher the proportion of relationships enhanced by structural holes, the more likely and able the entrepreneurial player, and so the more likely it is that the player's investments are in high-yield relationships. The result is a higher aggregate rate of return on investments.

I have shaded the area in the graph to indicate how I expect data to be distributed around the line of association. There is no imperative that says players have to take advantage of the benefits provided by structural holes. Players rich in entrepreneurial opportunity may choose to develop opportunities (and so appear in the upper right corner of the graph) or ignore them (and so appear in the lower right corner of the graph). Some players in Figure 1.7 are above the line. Some are below. If players

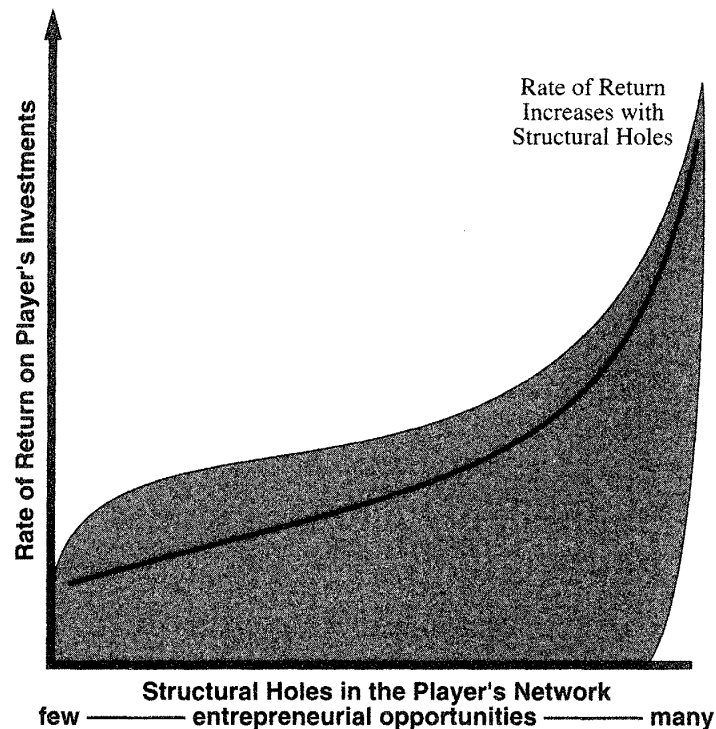


Figure 1.7 Rate of return and structural holes

were perfectly rational, observations would be clustered around the line. Players would take advantage of any entrepreneurial opportunity presented to them. A control for differences in player motivation, such as a McClelland measure of need for achievement, would have the same effect. The point is not the degree of deviation from the line of association; it is the greater deviation below the line. Variable motivation creates deviations below the true hole effect on rate of return.

This emphasizes the relative importance for empirical research of deviations above and below the line of association. Observations in the lower right corner of the graph, players underutilizing their entrepreneurial opportunities, might be due to variation in motivation. Observations in the upper left corner are a severe test of the argument. Players who have opportunities can choose whether to develop them. Players without opportunities do not have that choice. Within the limits of measurement error, there should be no observations in the upper left corner of the graph.

Secondary Holes

This brings me to the third component in the negotiation: the social structural conditions that constitute entrepreneurial opportunity. I have linked opportunity to structural holes, but not with respect to the whole domain of relevant holes. Thus far, a network optimized for entrepreneurial opportunity has a vine-and-cluster structure. As illustrated in Figures 1.3 and 1.4, a player has direct relations with primary contacts, each a port of access to a cluster of redundant secondary contacts. Structural holes between the primary contacts, a primary structural holes, provide information and control benefits. But the benefits they provide are affected by structural holes just beyond the border of the network. Structural holes among the secondary contacts within the cluster around each primary contact play a role in the *tertius* strategies. These are secondary structural holes.

CONTROL BENEFITS AND SECONDARY STRUCTURAL HOLES

The ultimate threat in negotiating a relationship is withdrawal: either severing your link to a former contact's cluster or transferring the primary relationship to a new person in the cluster. This threat depends on two things. First, there must be alternatives, secondary contacts who are redundant with your primary contact and capable of replacing the primary contact in your network. Examples include an alternative spouse in the case of negotiating a conjugal relationship, an alternative job in the case

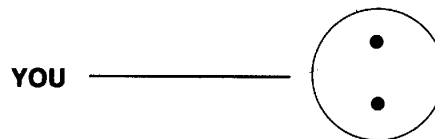
of negotiating with a truculent supervisor, or an alternative supplier in the case of a firm renewing a contract with a past supplier. Second, there must be structural holes among the secondary contacts. If there are no contacts substitutable for your current primary contact, he or she is free to impose demands—up to the limit of structural holes between primary contacts. If your current primary contact is in collusion with whatever substitutes exist, which eliminates structural holes you might exploit, he or she is free to impose demands—again, up to the limit of structural holes between primary contacts.

Consider Figure 1.8. You are negotiating with a primary contact in a cluster of redundant contacts indicated by dots enclosed by a gray circle. Situation A illustrates the familiar negotiation between buyer and seller. You use the offer from one buyer to raise the other's offer.

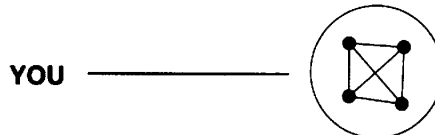
Situation B illustrates the exact opposite condition. Here the redundant contacts are all connected by strong relations. This is the situation of negotiating with a member of a social clique or cult. In the absence of holes over which you can broker the connection between redundant contacts, your only recourse is to live with your contact's demands, dominate the cluster, or cut the cluster from your network.

Network density is not the issue here. Situation C is a relatively low

A. Market



B. Clique



C. Hierarchy

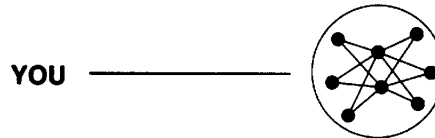


Figure 1.8 Contact clusters with and without secondary structural holes

density cluster (43% of the 28 relations within the cluster are marked with a line as strong), but contacts within the cluster are coordinated through their joint ties to two leaders in the center. It doesn't make sense to negotiate the price of a purchase in a department store by playing one sales clerk against another. They both answer to a higher authority. You have to make a purchase sufficiently large that it allows you to deal with someone higher in the organization. Then, as in Situation C, you can develop the structural hole between the two leaders at the center of the circle and play one leader against the other.

CLUSTER BOUNDARIES

Secondary contacts are a cluster of redundant players in the competitive arena beyond any one player's network. Players in the cluster are redundant by cohesion (strongly connected within the cluster) or structural equivalence (connected with the same players beyond the cluster). Given redundancy within clusters, the more general statement is that players are redundant contacts in the same cluster to the extent that they are connected with the same clusters of redundant contacts.

The idea is illustrated in Figure 1.9. Four identical networks are displayed at the top of the figure. Lines are relations, each gray circle indicates a cluster of redundant contacts, and the dark circle at the center is the player responsible for the network. Each network includes a primary contact in each of the six clusters.

The four central players are redundant. They are connected to the same clusters of redundant contacts, and so have the same information and control benefits. They might be connected to different people in each cluster, but their contacts are ports of access into the same six clusters. Rather than representing the four players with separate networks, it is more accurate to represent them as four redundant contacts within the dark circle in the network at the bottom of Figure 1.9. Contacts are aggregated similarly within each of the clusters identified by letter.

The same comparison illustrates nonredundancy. Notice the two bold curved lines between the player and two clusters, B and E. Clusters B and E are rich in structural holes, so relations with any contact in them will be more easily negotiated than relations with the better-organized clusters, such as A, D and F. Suppose that one of the central players decides to focus on these relations, leaving the other three to deal with clusters A, C, D, and E. The three are then no longer redundant with the first. The first is connected to clusters different from the ones in their networks.

This image of redundancy is analogous to the concept of substitutable producers in input-output economics. Two producers are substitutable in an economic network to the extent that they purchase similar volumes of the same kinds of supplies to make the commodity they sell. Suppliers are in turn substitutable to the extent that their product requires similar volumes of the same kinds of supplies. Two bakers are substitutable to the extent that they use the same kinds of ingredients. They might purchase their flour and sugar from different vendors, but they are substitut-

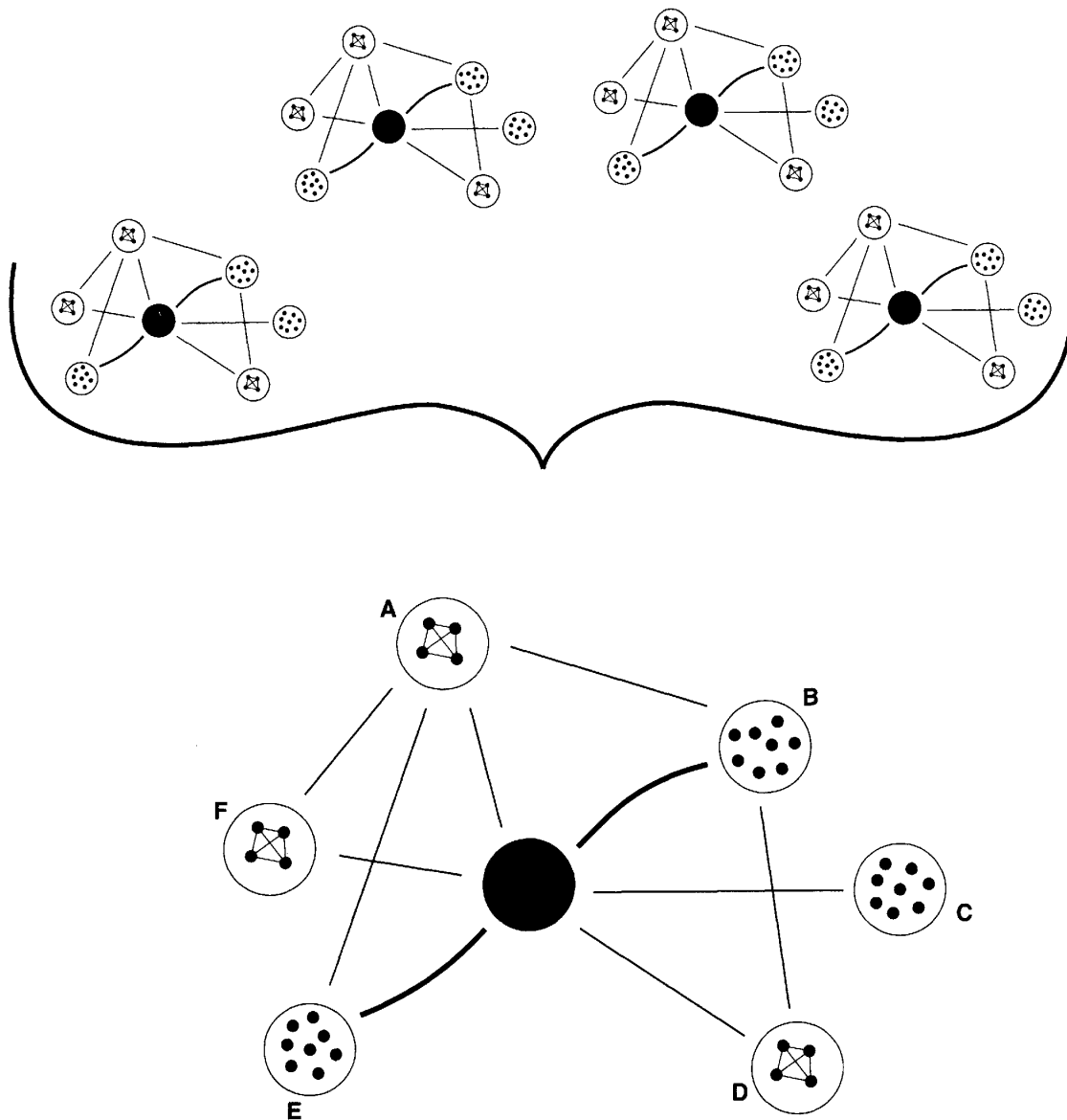


Figure 1.9 Four redundant networks pooled as one network surrounding four substitutable players

able to the extent that they similarly purchase flour from one of the alternative flour vendors and sugar from one of the alternative sugar vendors. Two auto manufacturers are substitutable to the extent that they use the same proportions of metal, glass, rubber, and plastic to produce the cars they sell. Each manufacturer might purchase glass from a different vendor, but they similarly purchase glass from one or another of the available glass vendors to make their cars. I will return to this point in the analysis of markets in Chapter 3 (Figure 3.2).

Redundancy as substitutability is analogous to the equivalence concept in network analysis, but different from the often-used variations of structural and role equivalence. Structurally equivalent people have identical relations with the same people. This is too narrow a definition of redundancy. The dark circles in the four networks in Figure 1.9 can have relations with completely different people within each cluster, which would make them redundant, but not structurally equivalent. At the other extreme, role equivalent people have identically structured relations, regardless of the specific individuals with whom they have relations. This is too broad a definition of redundancy. For example, a person connected only to cluster A in Figure 1.9 would be role equivalent to a person only connected to clusters D and F. They would be role equivalent in the sense of being an outsider connected to a clique; however, they are non-redundant because they are connected with different clusters of redundant contacts. Operationally, I am left with cluster boundaries defined a priori by some criterion, as I will illustrate in Chapter 3.

THE DEPTH OF A STRUCTURAL HOLE

Secondary only refers to the remove of a hole from the central player. Primary holes are between a player's direct contacts, secondary holes between indirect contacts. Of the two kinds of holes, the latter are the more intense.

Let the depth of a structural hole be the ease with which it can be developed for control and information benefits. When the hole is deep between two individuals, it is easy to play them against one another with *tertius* strategies.

Depth is characterized in Table 1.2 with combinations of the two indicators of holes: cohesion and equivalence. The columns contrast players who have no relationship with one another with players who meet frequently and feel emotionally close to one another (in other words, have a strong relationship). The rows contrast players in completely separate clusters with those who have equivalent ties to the same clusters (in other words, are close together in the same cluster).

Much of the table is clear from the cohesion and structural equivalence indicators for defining structural holes already discussed. There is a structural hole of some depth between the players in all conditions except the “no hole” cell. Redundancy is most likely between structurally equivalent people connected by a strong relationship. At the other extreme, there is a structural hole where both indicators show no connection: the “hole” cell in the upper left of Table 1.2. Redundancy is unlikely between total strangers in distant clusters.

Cohesion is a good indicator. Where cohesion is low, there is a hole between the players. There is no hole where cohesion is high between players equivalently connected to the same clusters. There is also a hole between players in distant clusters connected by a strong relationship. The two players are ports into different clusters of information, but their strong tie means a strong flow of information between them. Playing them against one another turns on the extent to which their cluster interests override their commitment to each other.

Cohesion is an especially good indicator relative to equivalence. The first row of the table shows a hole between players in separate clusters. But the second row shows that the widest extremes of hole depth occur between players in the same cluster. The second row of the table is the usual axis of imperfect market competition. Players connected to the same clusters are redundant, and so could replace one another in their respective networks. What I bring to your network, a contact connected to the same clusters that I reach could also bring to your network. I and the contact are substitutable producers; we are competitors in the same market. If I have strong relations with my colleagues, we collude to avoid people playing us against one another, and you face a cluster like the one in Figure 1.8B. If the relations are poor among my colleagues and myself, we are easy prey to being played against one another because we are so readily substitutable, and you face a cluster like the one in Figure 1.8A.

Equivalence is the frame and cohesion the indicator. Equivalent ties to the same clusters frame two players as competitors in the same market.

Table 1.2 Depth of a structural hole between players

| Equivalent ties to clusters | Cohesion between players | |
|-----------------------------|--------------------------|---------|
| | None | Strong |
| None | HOLE | SHALLOW |
| Strong | DEEP | NO HOLE |

Cohesion defines the depth of the hole between them. In terms of a regression model, the depth of the hole between two players increases with their equivalence, decreases with the strength of relation between them, and decreases sharply with the extent to which they are equivalent and strongly connected.

Structural Autonomy

The argument can now be summarized with a concept defining the extent to which a player's network is rich in structural holes, and thus rich in entrepreneurial opportunity, and thus rich in information and control benefits. The concept is structural autonomy. I will present the concept in a general way here, postponing detailed discussion for Chapter 2.

The argument began with a generic production equation. Profit equals an investment multiplied by a rate of return. The benefits of a relationship can be expressed in an analogous form: time and energy invested to reach a contact multiplied by a rate of return. A player's entrepreneurial opportunities are enhanced by a relationship to the extent that: (a) the player has invested substantial time and energy to secure a connection with the contact, and (b) there are many structural holes around the contact ensuring a high rate of return on the investment. More specifically, rate of return concerns how and whom you reach with the relationship. Time and energy invested to reach a player with more resources generates more social capital. For the sake of argument, as explained in the discussion of social capital, I assume that a player with a network optimized for structural holes can identify suitably endowed contacts. My concern is the how of a relationship, defined by the structure of a network and its connection with the social structure of the competitive arena. Thus the rate of return keyed to structural holes is a product of the extent to which there are: (a) many primary structural holes between the contact and others in the player's network, and (b) many secondary structural holes between the contact and others outside the network who could replace the contact.

There is also the issue of structural holes around the player. As the holes around contacts provide information and control benefits to the player, holes around the player can be developed by contacts for their benefit. Consider your position as one of four disconnected players at the center of the network at the bottom of Figure 1.9. Your contacts have the option of replacing you with one of your colleagues who provides the same network benefits that you do. To manage this uncertainty, you

might develop relationships with your colleagues so that it would be difficult to play them off against you (an oligopoly strategy), or you might specialize in some way so that they no longer provide network benefits redundant with your own (a differentiation strategy). The issue of strategic response is the subject of Chapter 7. The point here is that your negotiating position is weaker than expected from the distribution of structural holes around contacts. Developing entrepreneurial opportunities depends on having numerous structural holes around your contacts and none attached to yourself.

These considerations come together in the concept of structural autonomy. Players with relationships free of structural holes at their own end and rich in structural holes at the other end are structurally autonomous. These are the players best positioned for the information and control benefits that a network can provide. These are the players to the far right of the graph in Figure 1.7. Structural autonomy summarizes the action potential of the *tertius's* network. The budget equation for optimizing structural autonomy has an upper limit set by the time and energy of the *tertius*, and a trade-off between the structural holes a new contact provides versus the time and energy required to maintain a productive relationship with the contact.¹⁴

Summary

This chapter contains the core argument of the book. Structural holes are introduced and the manner in which they are a competitive advantage is explained.

The argument begins with the task of profit. Profit is generated by a production equation in which player investments are multiplied by the going rate of return. A million dollars invested at a 10 percent rate of return yields a hundred thousand dollar profit. Investments create an ability to produce a competitive product. Capital is invested, for example, to build and operate a factory. Rate of return is an opportunity to profit from the investment.

The rate of return is keyed to the social structure of the competitive arena. Each player has a network of contacts in the competitive arena. Certain players are connected to certain others, trusting of certain others, obligated to support certain others, dependent on exchange with certain others. Something about the structure of the player's network and the location of the player's contacts in the social structure of the arena defines the player's chances of getting higher rates of return on investment.

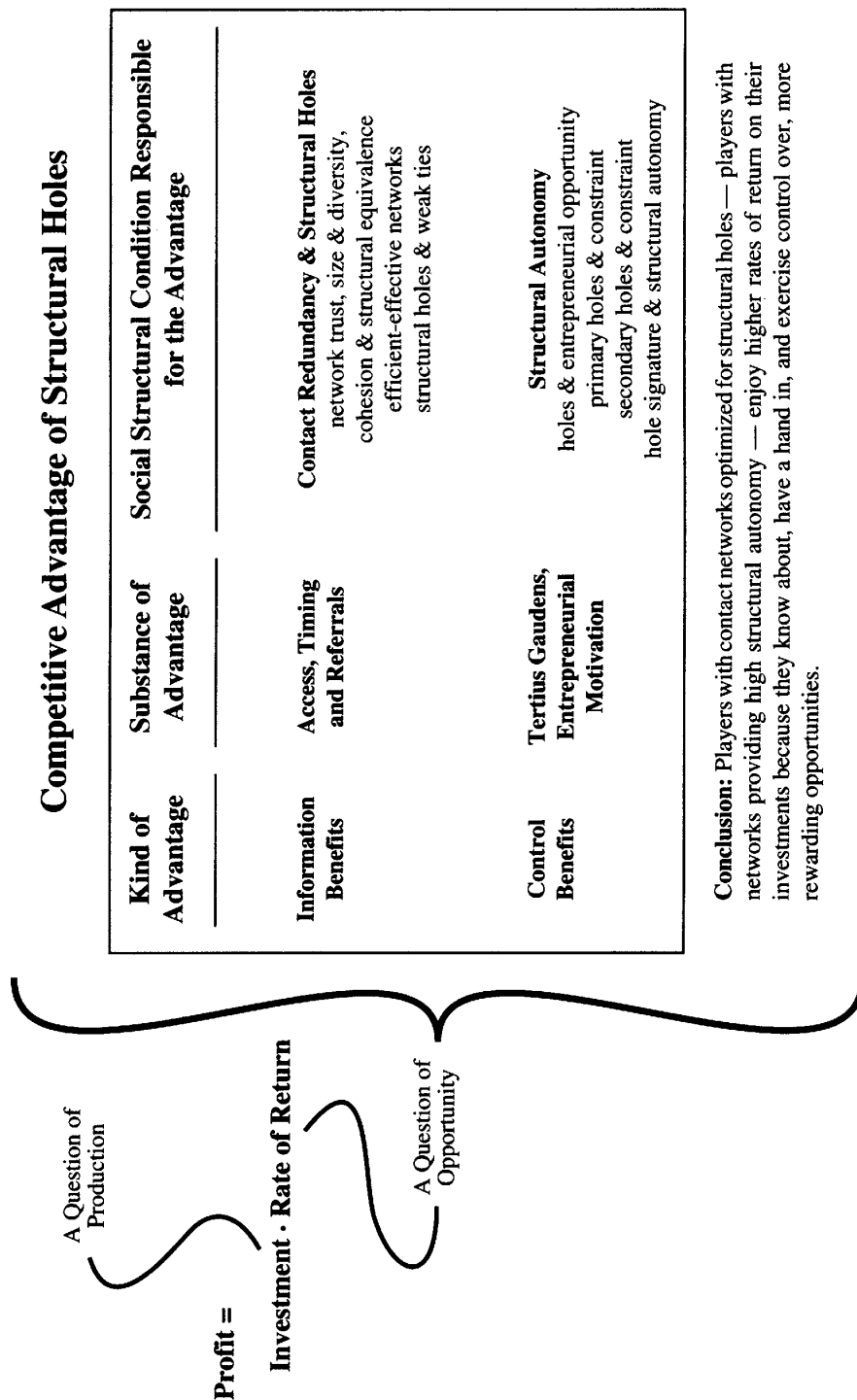


Figure 1.10 Argument

The chances are enhanced by two kinds of network benefits, information and control, distinguished by the rows of the box in Figure 1.10.

The substance of information benefits are access, timing, and referrals. The player's network provides access to information well beyond what the player could process alone. The network also provides that information early, which gives the player an advantage in acting on the information. These benefits concern information coming to the player from contacts. Referral benefits involve the opposite flow. The network that filters information coming to the player also directs, concentrates, and legitimates information received by others about the player. Referrals get the player's interests represented in a positive light, at the right time, in the right places.

Information benefits are maximized in a large, diverse network of trusted contacts. Trust is important with respect to the honoring of interpersonal debt by contacts, but is an idiographic question answered by the social match between player and each contact individually. Network size and diversity under a presumption of trust are the general parameters to be optimized. The effective size of a network can be less than its observed size. Size is the number of primary contacts in a network; effective size is the number of nonredundant contacts. Two contacts are redundant to the extent that they provide the same information benefits to the player. Cohesion is an empirical indicator of redundancy. Contacts strongly connected to each other are likely to have similar information and so provide redundant benefit to the player. Structural equivalence is a second indicator. Contacts who, regardless of their relationship with one another, link the player to the same third parties have the same sources of information, and so provide redundant benefit to the player. Structural holes are the gaps between nonredundant contacts. As a result of the hole between them, the two contacts provide network benefits that are in some degree additive rather than overlapping. A network optimized for information benefits can be described with respect to its contacts or its connections between contacts. A network rich in nonredundant contacts is rich in structural holes.

The structural holes that generate information benefits also generate control benefits, giving certain players an advantage in negotiating their relationships. Sociological theory offers a role describing people who derive control benefits from structural holes. It is the *tertius gaudens*, the third who benefits: a person who derives benefit from brokering relationships between other players. There are two *tertius* strategies. People can be played against one another when they compete for the same rela-

tionship: for example, two buyers after the same purchase. Second, people can be played against one another when they make conflicting demands on the same individual in separate relationships: a science professor's course demands, for example, being played by a student against the course demands of a humanities professor. There is a presumption of tension here, but the essential tension is not the hostility of combatants; it is merely uncertainty. Separating the uncertainty of control from its consequences, *tertius* strategies apply similarly to negotiating control that has severe consequences or to negotiating control that is of little consequence. What is essential is that the control is uncertain, that no one can act as if he or she has absolute authority in the relationship under negotiation. In the swirling mix of preferences characteristic to social networks, where no demands have absolute authority, the *tertius* negotiates for favorable terms.

The information and control benefits are multiplicative, augmenting and dependent on one another, together emerging from the wellspring of structural holes in a network. Structural holes are the setting for *tertius* strategies. Information is the substance. Accurate, ambiguous, or distorted information is moved between contacts by the *tertius*. One bidder is informed of a competitive offer in the first strategy. A player in one relationship is informed of demands from other relationships in the second strategy.

The final task is to bring the argument together in a definition, relevant to empirical research, of the extent to which a player's network is rich in structural holes, and so entrepreneurial opportunity, and so information and control benefits. Each of a player's relationships is treated as an investment on which structural holes determine the rate of return. A player's entrepreneurial opportunities are enhanced by a relationship to the extent that: (a) the player has invested substantial time and energy to secure a connection with the contact, and (b) there are many structural holes around the contact ensuring a high rate of return on the investment. The rate of return keyed to structural holes is a product of the extent to which there are: (a) many primary structural holes between the contact and others in the player's network, and (b) many secondary structural holes between the contact and others outside the network who could replace the contact. There are also the structural holes around the player. As the holes around contacts provide information and control benefits to the player, holes around the player can be developed by contacts for their benefit.

These considerations come together in the concept of structural auton-

omy. Players with relationships free of structural holes at their own end and rich in structural holes at the other end are structurally autonomous. These are the players best positioned for the information and control benefits that a network can provide. Structural autonomy summarizes the action potential of the *tertius*'s network. The budget equation for optimizing structural autonomy has an upper limit set by the *tertius*'s time and energy, and a trade-off between the structural holes a new contact provides versus the time and energy required to maintain a productive relationship with the contact. The summary conclusion is that players with networks optimized for structural holes—players with networks providing high structural autonomy—enjoy higher rates of return on their investments because they know about, have a hand in, and exercise control over, more rewarding opportunities.

The reasoning isn't new. The argument draws on social psychological studies of negotiation, economic studies of imperfect competition, and, most especially, sociological studies of roles and statuses in social structure.

What is new is the expression of competitive advantage—in economic, political, or social arenas—in terms of structural holes as an elemental unit clearly defined in theory and readily operationalized for empirical research.