

# Networks, Crowds, and Markets:

## Reasoning about a Highly Connected World

David Easley  
Dept. of Economics  
Cornell University

Jon Kleinberg  
Dept. of Computer Science  
Cornell University

Cambridge University Press, 2010  
Draft version: June 10, 2010.



# Contents

<b>Preface</b>	<b>i</b>
<b>1 Overview</b>	<b>1</b>
1.1 Aspects of Networks . . . . .	2
1.2 Central Themes and Topics . . . . .	8
<b>I Graph Theory and Social Networks</b>	<b>21</b>
<b>2 Graphs</b>	<b>23</b>
2.1 Basic Definitions . . . . .	23
2.2 Paths and Connectivity . . . . .	25
2.3 Distance and Breadth-First Search . . . . .	32
2.4 Network Datasets: An Overview . . . . .	40
2.5 Exercises . . . . .	44
<b>3 Strong and Weak Ties</b>	<b>47</b>
3.1 Triadic Closure . . . . .	48
3.2 The Strength of Weak Ties . . . . .	50
3.3 Tie Strength and Network Structure in Large-Scale Data . . . . .	56
3.4 Tie Strength, Social Media, and Passive Engagement . . . . .	60
3.5 Closure, Structural Holes, and Social Capital . . . . .	64
3.6 Advanced Material: Betweenness Measures and Graph Partitioning . . . . .	69
3.7 Exercises . . . . .	83
<b>4 Networks in Their Surrounding Contexts</b>	<b>85</b>
4.1 Homophily . . . . .	86
4.2 Mechanisms Underlying Homophily: Selection and Social Influence . . . . .	90
4.3 Affiliation . . . . .	93
4.4 Tracking Link Formation in On-Line Data . . . . .	97
4.5 A Spatial Model of Segregation . . . . .	107
4.6 Exercises . . . . .	116

<b>5</b>	<b>Positive and Negative Relationships</b>	<b>119</b>
5.1	Structural Balance . . . . .	120
5.2	Characterizing the Structure of Balanced Networks . . . . .	123
5.3	Applications of Structural Balance . . . . .	126
5.4	A Weaker Form of Structural Balance . . . . .	129
5.5	Advanced Material: Generalizing the Definition of Structural Balance . . . . .	132
5.6	Exercises . . . . .	148
<b>II</b>	<b>Game Theory</b>	<b>153</b>
<b>6</b>	<b>Games</b>	<b>155</b>
6.1	What is a Game? . . . . .	156
6.2	Reasoning about Behavior in a Game . . . . .	158
6.3	Best Responses and Dominant Strategies . . . . .	163
6.4	Nash Equilibrium . . . . .	166
6.5	Multiple Equilibria: Coordination Games . . . . .	168
6.6	Multiple Equilibria: The Hawk-Dove Game . . . . .	172
6.7	Mixed Strategies . . . . .	173
6.8	Mixed Strategies: Examples and Empirical Analysis . . . . .	179
6.9	Pareto-Optimality and Social Optimality . . . . .	184
6.10	Advanced Material: Dominated Strategies and Dynamic Games . . . . .	186
6.11	Exercises . . . . .	200
<b>7</b>	<b>Evolutionary Game Theory</b>	<b>209</b>
7.1	Fitness as a Result of Interaction . . . . .	210
7.2	Evolutionarily Stable Strategies . . . . .	211
7.3	A General Description of Evolutionarily Stable Strategies . . . . .	216
7.4	Relationship Between Evolutionary and Nash Equilibria . . . . .	218
7.5	Evolutionarily Stable Mixed Strategies . . . . .	220
7.6	Exercises . . . . .	225
<b>8</b>	<b>Modeling Network Traffic using Game Theory</b>	<b>229</b>
8.1	Traffic at Equilibrium . . . . .	229
8.2	Braess's Paradox . . . . .	231
8.3	Advanced Material: The Social Cost of Traffic at Equilibrium . . . . .	233
8.4	Exercises . . . . .	243
<b>9</b>	<b>Auctions</b>	<b>249</b>
9.1	Types of Auctions . . . . .	249
9.2	When are Auctions Appropriate? . . . . .	251
9.3	Relationships between Different Auction Formats . . . . .	252
9.4	Second-Price Auctions . . . . .	254
9.5	First-Price Auctions and Other Formats . . . . .	257
9.6	Common Values and The Winner's Curse . . . . .	258

9.7	Advanced Material: Bidding Strategies in First-Price and All-Pay Auctions . . . . .	260
9.8	Exercises . . . . .	268

### **III Markets and Strategic Interaction in Networks 275**

#### **10 Matching Markets 277**

10.1	Bipartite Graphs and Perfect Matchings . . . . .	277
10.2	Valuations and Optimal Assignments . . . . .	282
10.3	Prices and the Market-Clearing Property . . . . .	284
10.4	Constructing a Set of Market-Clearing Prices . . . . .	288
10.5	How Does this Relate to Single-Item Auctions? . . . . .	291
10.6	Advanced Material: A Proof of the Matching Theorem . . . . .	293
10.7	Exercises . . . . .	302

#### **11 Network Models of Markets with Intermediaries 311**

11.1	Price-Setting in Markets . . . . .	311
11.2	A Model of Trade on Networks . . . . .	315
11.3	Equilibria in Trading Networks . . . . .	322
11.4	Further Equilibrium Phenomena: Auctions and Ripple Effects . . . . .	326
11.5	Social Welfare in Trading Networks . . . . .	330
11.6	Trader Profits . . . . .	332
11.7	Reflections on Trade with Intermediaries . . . . .	334
11.8	Exercises . . . . .	334

#### **12 Bargaining and Power in Networks 339**

12.1	Power in Social Networks . . . . .	339
12.2	Experimental Studies of Power and Exchange . . . . .	342
12.3	Results of Network Exchange Experiments . . . . .	344
12.4	A Connection to Buyer-Seller Networks . . . . .	348
12.5	Modeling Two-Person Interaction: The Nash Bargaining Solution . . . . .	349
12.6	Modeling Two-Person Interaction: The Ultimatum Game . . . . .	352
12.7	Modeling Network Exchange: Stable Outcomes . . . . .	355
12.8	Modeling Network Exchange: Balanced Outcomes . . . . .	359
12.9	Advanced Material: A Game-Theoretic Approach to Bargaining . . . . .	361
12.10	Exercises . . . . .	369

### **IV Information Networks and the World Wide Web 373**

#### **13 The Structure of the Web 375**

13.1	The World Wide Web . . . . .	376
13.2	Information Networks, Hypertext, and Associative Memory . . . . .	378
13.3	The Web as a Directed Graph . . . . .	384
13.4	The Bow-Tie Structure of the Web . . . . .	388

13.5	The Emergence of Web 2.0 . . . . .	392
13.6	Exercises . . . . .	394
<b>14</b>	<b>Link Analysis and Web Search</b>	<b>397</b>
14.1	Searching the Web: The Problem of Ranking . . . . .	397
14.2	Link Analysis using Hubs and Authorities . . . . .	399
14.3	PageRank . . . . .	406
14.4	Applying Link Analysis in Modern Web Search . . . . .	412
14.5	Applications beyond the Web . . . . .	415
14.6	Advanced Material: Spectral Analysis, Random Walks, and Web Search . . .	417
14.7	Exercises . . . . .	429
<b>15</b>	<b>Sponsored Search Markets</b>	<b>437</b>
15.1	Advertising Tied to Search Behavior . . . . .	437
15.2	Advertising as a Matching Market . . . . .	440
15.3	Encouraging Truthful Bidding in Matching Markets: The VCG Principle . .	444
15.4	Analyzing the VCG Procedure: Truth-Telling as a Dominant Strategy . . . .	449
15.5	The Generalized Second Price Auction . . . . .	452
15.6	Equilibria of the Generalized Second Price Auction . . . . .	456
15.7	Ad Quality . . . . .	459
15.8	Complex Queries and Interactions Among Keywords . . . . .	461
15.9	Advanced Material: VCG Prices and the Market-Clearing Property . . . . .	462
15.10	Exercises . . . . .	478
<b>V</b>	<b>Network Dynamics: Population Models</b>	<b>481</b>
<b>16</b>	<b>Information Cascades</b>	<b>483</b>
16.1	Following the Crowd . . . . .	483
16.2	A Simple Herding Experiment . . . . .	485
16.3	Bayes' Rule: A Model of Decision-Making Under Uncertainty . . . . .	489
16.4	Bayes' Rule in the Herding Experiment . . . . .	494
16.5	A Simple, General Cascade Model . . . . .	496
16.6	Sequential Decision-Making and Cascades . . . . .	500
16.7	Lessons from Cascades . . . . .	503
16.8	Exercises . . . . .	505
<b>17</b>	<b>Network Effects</b>	<b>509</b>
17.1	The Economy Without Network Effects . . . . .	510
17.2	The Economy with Network Effects . . . . .	514
17.3	Stability, Instability, and Tipping Points . . . . .	517
17.4	A Dynamic View of the Market . . . . .	519
17.5	Industries with Network Goods . . . . .	526
17.6	Mixing Individual Effects with Population-Level Effects . . . . .	528
17.7	Advanced Material: Negative Externalities and The El Farol Bar Problem .	533

17.8 Exercises . . . . .	541
<b>18 Power Laws and Rich-Get-Richer Phenomena</b>	<b>543</b>
18.1 Popularity as a Network Phenomenon . . . . .	543
18.2 Power Laws . . . . .	545
18.3 Rich-Get-Richer Models . . . . .	547
18.4 The Unpredictability of Rich-Get-Richer Effects . . . . .	549
18.5 The Long Tail . . . . .	551
18.6 The Effect of Search Tools and Recommendation Systems . . . . .	554
18.7 Advanced Material: Analysis of Rich-Get-Richer Processes . . . . .	555
18.8 Exercises . . . . .	559
 <b>VI Network Dynamics: Structural Models</b>	 <b>561</b>
<b>19 Cascading Behavior in Networks</b>	<b>563</b>
19.1 Diffusion in Networks . . . . .	563
19.2 Modeling Diffusion through a Network . . . . .	565
19.3 Cascades and Clusters . . . . .	573
19.4 Diffusion, Thresholds, and the Role of Weak Ties . . . . .	578
19.5 Extensions of the Basic Cascade Model . . . . .	580
19.6 Knowledge, Thresholds, and Collective Action . . . . .	583
19.7 Advanced Material: The Cascade Capacity . . . . .	587
19.8 Exercises . . . . .	603
 <b>20 The Small-World Phenomenon</b>	 <b>611</b>
20.1 Six Degrees of Separation . . . . .	611
20.2 Structure and Randomness . . . . .	612
20.3 Decentralized Search . . . . .	616
20.4 Modeling the Process of Decentralized Search . . . . .	619
20.5 Empirical Analysis and Generalized Models . . . . .	622
20.6 Core-Periphery Structures and Difficulties in Decentralized Search . . . . .	629
20.7 Advanced Material: Analysis of Decentralized Search . . . . .	631
20.8 Exercises . . . . .	642
 <b>21 Epidemics</b>	 <b>645</b>
21.1 Diseases and the Networks that Transmit Them . . . . .	645
21.2 Branching Processes . . . . .	647
21.3 The SIR Epidemic Model . . . . .	650
21.4 The SIS Epidemic Model . . . . .	656
21.5 Synchronization . . . . .	659
21.6 Transient Contacts and the Dangers of Concurrency . . . . .	662
21.7 Genealogy, Genetic Inheritance, and Mitochondrial Eve . . . . .	666
21.8 Advanced Material: Analysis of Branching and Coalescent Processes . . . . .	672
21.9 Exercises . . . . .	685

<b>VII</b>	<b>Institutions and Aggregate Behavior</b>	<b>689</b>
<b>22</b>	<b>Markets and Information</b>	<b>691</b>
22.1	Markets with Exogenous Events . . . . .	692
22.2	Horse Races, Betting, and Beliefs . . . . .	694
22.3	Aggregate Beliefs and the “Wisdom of Crowds” . . . . .	700
22.4	Prediction Markets and Stock Markets . . . . .	704
22.5	Markets with Endogenous Events . . . . .	708
22.6	The Market for Lemons . . . . .	709
22.7	Asymmetric Information in Other Markets . . . . .	714
22.8	Signaling Quality . . . . .	718
22.9	Quality Uncertainty On-Line: Reputation Systems and Other Mechanisms . . . . .	720
22.10	Advanced Material: Wealth Dynamics in Markets . . . . .	723
22.11	Exercises . . . . .	730
<b>23</b>	<b>Voting</b>	<b>735</b>
23.1	Voting for Group Decision-Making . . . . .	735
23.2	Individual Preferences . . . . .	737
23.3	Voting Systems: Majority Rule . . . . .	740
23.4	Voting Systems: Positional Voting . . . . .	745
23.5	Arrow’s Impossibility Theorem . . . . .	748
23.6	Single-Peaked Preferences and the Median Voter Theorem . . . . .	750
23.7	Voting as a Form of Information Aggregation . . . . .	756
23.8	Insincere Voting for Information Aggregation . . . . .	758
23.9	Jury Decisions and the Unanimity Rule . . . . .	761
23.10	Sequential Voting and the Relation to Information Cascades . . . . .	766
23.11	Advanced Material: A Proof of Arrow’s Impossibility Theorem . . . . .	767
23.12	Exercises . . . . .	772
<b>24</b>	<b>Property Rights</b>	<b>775</b>
24.1	Externalities and the Coase Theorem . . . . .	775
24.2	The Tragedy of the Commons . . . . .	780
24.3	Intellectual Property . . . . .	783
24.4	Exercises . . . . .	786



# Chapter 20

## The Small-World Phenomenon

### 20.1 Six Degrees of Separation

In the previous chapter, we considered how social networks can serve as conduits by which ideas and innovations flow through groups of people. To develop this idea more fully, we now relate it to another basic structural issue — the fact that these groups can be connected by very short paths through the social network. When people try to use these short paths to reach others who are socially distant, they are engaging in a kind of “focused” search that is much more targeted than the broad spreading pattern exhibited by the diffusion of information or a new behavior. Understanding the relationship between targeted search and wide-ranging diffusion is important in thinking more generally about the way things flow through social networks.

As we saw in Chapter 2, the fact that social networks are so rich in short paths is known as the *small-world phenomenon*, or the “six degrees of separation,” and it has long been the subject of both anecdotal and scientific fascination. To briefly recapitulate what we discussed in that earlier chapter, the first significant empirical study of the small-world phenomenon was undertaken by the social psychologist Stanley Milgram [297, 391], who asked randomly chosen “starter” individuals to each try forwarding a letter to a designated “target” person living in the town of Sharon, MA, a suburb of Boston. He provided the target’s name, address, occupation, and some personal information, but stipulated that the participants could not mail the letter directly to the target; rather, each participant could only advance the letter by forwarding it to a single acquaintance that he or she knew on a first-name basis, with the goal of reaching the target as rapidly as possible. Roughly a third of the letters eventually arrived at the target, in a median of six steps, and this has since served as

---

D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010. Draft version: June 10, 2010.

basic experimental evidence for the existence of short paths in the global friendship network, linking all (or almost all) of us together in society. This style of experiment, constructing paths through social networks to distant target people, has been repeated by a number of other groups in subsequent decades [131, 178, 257].

Milgram’s experiment really demonstrated two striking facts about large social networks: first, that short paths are there in abundance; and second, that people, acting without any sort of global “map” of the network, are effective at collectively finding these short paths. It is easy to imagine a social network where the first of these is true but the second isn’t — a world where the short paths are there, but where a letter forwarded from thousands of miles away might simply wander from one acquaintance to another, lost in a maze of social connections [248]. A large social-networking site where everyone was known only by 9-digit pseudonyms would be like this: if you were told, “Forward this letter to user number 482285204, using only people you know on a first-name basis,” the task would clearly be hopeless. The real global friendship network contains enough clues about how people fit together in larger structures — both geographic and social — to allow the process of search to focus in on distant targets. Indeed, when Killworth and Bernard performed follow-up work on the Milgram experiment, studying the strategies that people employ for choosing how to forward a message toward a target, they found a mixture of primarily geographic and occupational features being used, with different features being favored depending on the characteristics of the target in relation to the sender [243].

We begin by developing models for both of these principles — the existence of short paths and also the fact that they can be found. We then look at how some of these models are borne out to a surprising extent on large-scale social-network data. Finally, in Section 20.6, we look at some of the fragility of the small-world phenomenon, and the caveats that must be considered in thinking about it: particularly the fact that people are most successful at finding paths when the target is high-status and socially accessible [255]. The picture implied by these difficulties raises interesting additional points about the global structure of social networks, and suggests questions for further research.

## 20.2 Structure and Randomness

Let’s start with models for the existence of short paths: Should we be surprised by the fact that the paths between seemingly arbitrary pairs of people are so short? Figure 20.1(a) illustrates a basic argument suggesting that short paths are at least compatible with intuition. Suppose each of us knows more than 100 other people on a first-name basis (in fact, for most people, the number is significantly larger). Then, taking into account the fact that each of your friends has at least 100 friends other than you, you could in principle be two steps away from over  $100 \cdot 100 = 10,000$  people. Taking into account the 100 friends of these

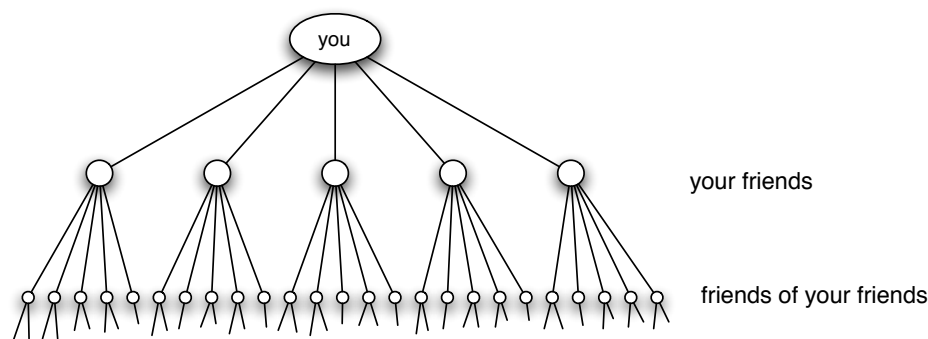
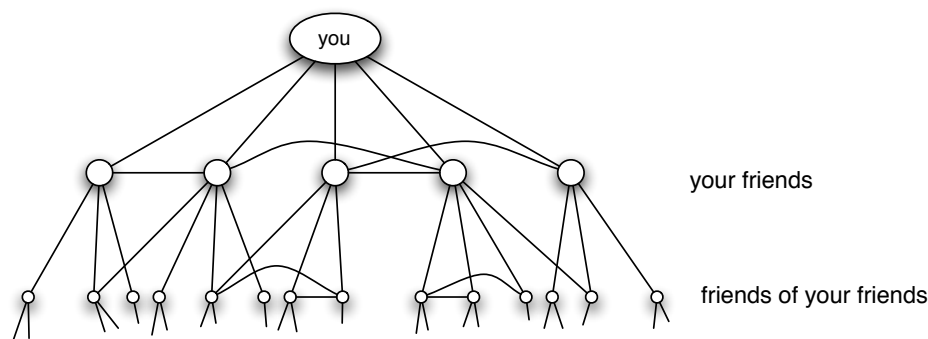
(a) *Pure exponential growth produces a small world*(b) *Triadic closure reduces the growth rate*

Figure 20.1: Social networks expand to reach many people in only a few steps.

people brings us to more than  $100 \cdot 100 \cdot 100 = 1,000,000$  people who in principle could be three steps away. In other words, the numbers are growing by powers of 100 with each step, bringing us to 100 million after four steps, and 10 billion after five steps.

There's nothing mathematically wrong with this reasoning, but it's not clear how much it tells us about real social networks. The difficulty already manifests itself with the second step, where we conclude that there may be more than 10,000 people within two steps of you. As we've seen, social networks abound in triangles — sets of three people who mutually know each other — and in particular, many of your 100 friends will know each other. As a result, when we think about the nodes you can reach by following edges from your friends, many of these edges go from one friend to another, not to the rest of world, as illustrated schematically in Figure 20.1(b). The number 10,000 came from assuming that each of your 100 friends was linked to 100 *new* people; and without this, the number of friends you could reach in two steps could be much smaller.

So the effect of triadic closure in social networks works to limit the number of people you can reach by following short paths, as shown by the contrast between Figures 20.1(a)

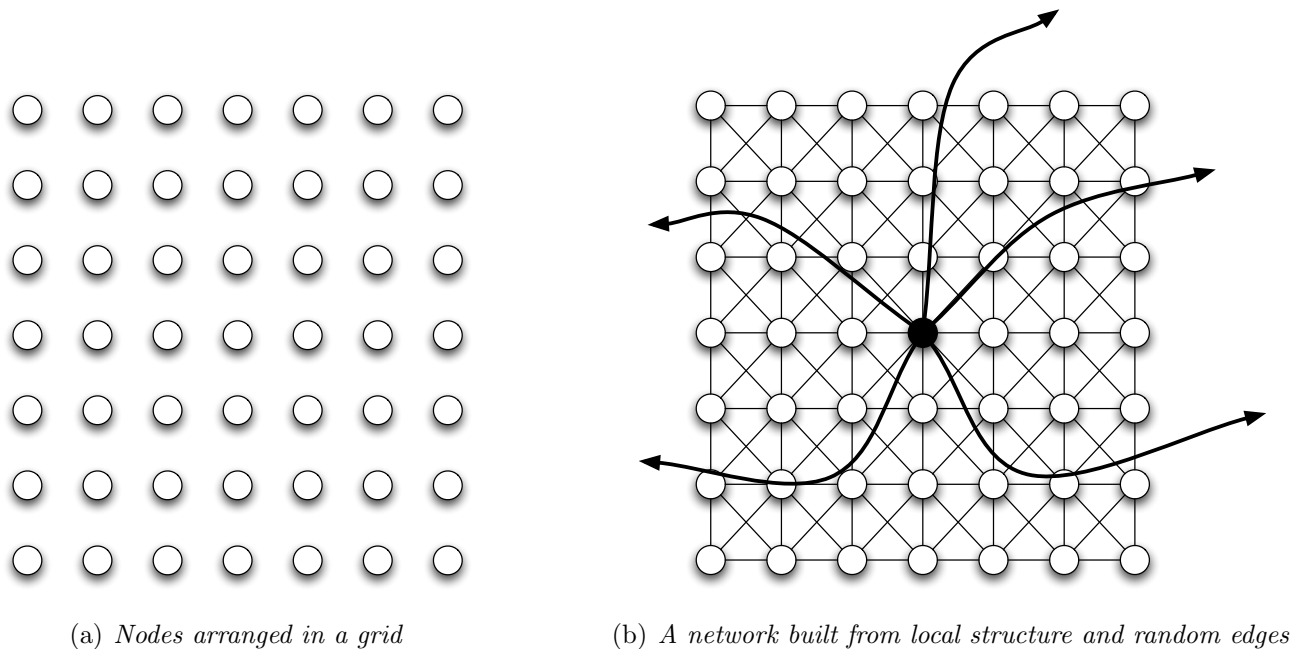


Figure 20.2: The Watts-Strogatz model arises from a highly clustered network (such as the grid), with a small number of random links added in.

and 20.1(b). And, indeed, at an implicit level, this is a large part of what makes the small-world phenomenon surprising to many people when they first hear it: the social network appears from the local perspective of any one individual to be highly clustered, not the kind of massively branching structure that would more obviously reach many nodes along very short paths.

**The Watts-Strogatz model.** Can we make up a simple model that exhibits both of the features we've been discussing: many closed triads, but also very short paths? In 1998, Duncan Watts and Steve Strogatz argued [411] that such a model follows naturally from a combination of two basic social-network ideas that we saw in Chapters 3 and 4: homophily (the principle that we connect to others who are like ourselves) and weak ties (the links to acquaintances that connect us to parts of the network that would otherwise be far away). Homophily creates many triangles, while the weak ties still produce the kind of widely branching structure that reaches many nodes in a few steps.

Watts and Strogatz made this proposal concrete in a very simple model that generates random networks with the desired properties. Paraphrasing their original formulation slightly (but keeping the main idea intact), let's suppose that everyone lives on a two-dimensional grid — we can imagine the grid as a model of geographic proximity, or potentially some more abstract kind of social proximity, but in any case a notion of similarity that guides the formation of links. Figure 20.2(a) shows the set of nodes arranged on a grid; we say that

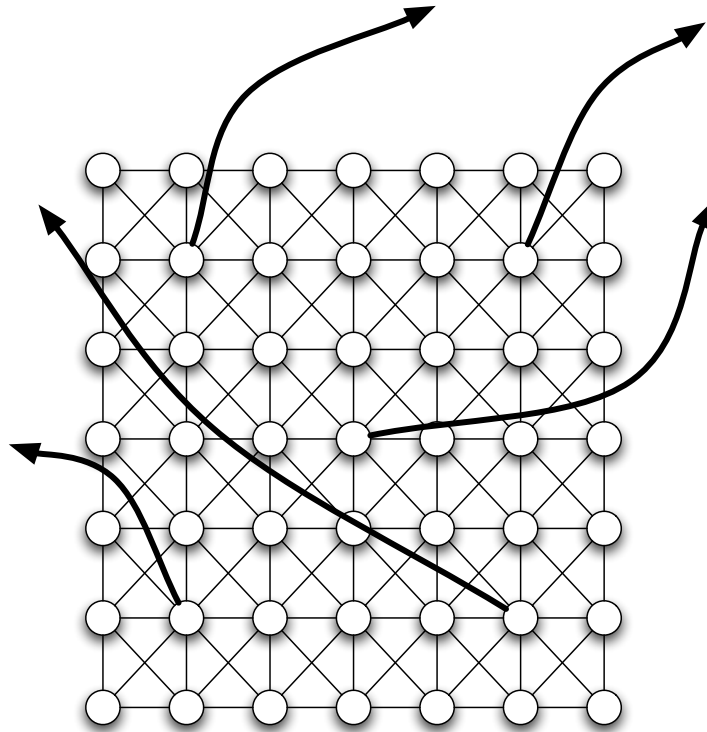


Figure 20.3: The general conclusions of the Watts-Strogatz model still follow even if only a small fraction of the nodes on the grid each have a *single* random link.

two nodes are one *grid step* apart if they are directly adjacent to each other in either the horizontal or vertical direction.

We now create a network by giving each node two kinds of links: those explainable purely by homophily, and those that constitute weak ties. Homophily is captured by having each node form a link to all other nodes that lie within a radius of up to  $r$  grid steps away, for some constant value of  $r$ : these are the links you form to people because you are similar to them. Then, for some other constant value  $k$ , each node also forms a link to  $k$  other nodes selected uniformly at random from the grid — these correspond to weak ties, connecting nodes who lie very far apart on the grid.

Figure 20.2(b) gives a schematic picture of the resulting network — a hybrid structure consisting of a small amount of randomness (the weak ties) sprinkled onto an underlying structured pattern (the homophilous links). Watts and Strogatz observe first that the network has many triangles: any two neighboring nodes (or nearby nodes) will have many common friends, where their neighborhoods of radius  $r$  overlap, and this produces many triangles. But they also find that there are — with high probability — very short paths connecting every pair of nodes in the network. Roughly, the argument is as follows. Suppose

we start tracing paths outward from a starting node  $v$ , using only the  $k$  random weak ties out of each node. Since these edges link to nodes chosen uniformly at random, we are very unlikely to ever see a node twice in the first few steps outward from  $v$ . As a result, these first few steps look almost like the picture in Figure 20.1(a), when there was no triadic closure, and so a huge number of nodes are reached in a small number of steps. A mathematically precise version of this argument was carried out by Bollobás and Chung [67], who determined the typical lengths of paths that it implies.

Once we understand how this type of hybrid network leads to short paths, we in fact find that a surprisingly small amount of randomness is needed to achieve the same qualitative effect. Suppose, for example, that instead of allowing each node to have  $k$  random friends, we only allow one out of every  $k$  nodes to have a *single* random friend — keeping the proximity-based edges as before, as illustrated schematically in Figure 20.3. Loosely speaking, we can think of this model with fewer random friends as corresponding to a technologically earlier time, when most people only knew their near neighbors, and a few people knew someone far away. Even this network will have short paths between all pairs of nodes. To see why, suppose that we conceptually group  $k \times k$  subsquares of the grid into “towns.” Now, consider the small-world phenomenon at the level of towns. Each town contains approximately  $k$  people who each have a random friend, and so the town collectively has  $k$  links to other towns selected uniformly at random. So this is just like the previous model, except that towns are now playing the role of individual nodes — and so we can find short paths between any pair of towns. But now to find a short path between any two people, we first find a short path between the two towns they inhabit, and then use the proximity-based edges to turn this into an actual path in the network on individual people.

This, then, is the crux of the Watts-Strogatz model: introducing a tiny amount of randomness — in the form of long-range weak ties — is enough to make the world “small,” with short paths between every pair of nodes.

## 20.3 Decentralized Search

Let’s now consider the second basic aspect of the Milgram small-world experiment — the fact that people were actually able to collectively find short paths to the designated target. This novel kind of “social search” task was a necessary consequence of the way Milgram formulated the experiment for his participants. To really find the *shortest* path from a starting person to the target, one would have to instruct the starter to forward a letter to *all* of his or her friends, who in turn should have forwarded the letter to all of their friends, and so forth. This “flooding” of the network would have reached the target as rapidly as possible — it is essentially the *breadth-first search* procedure from Chapter 2 — but for obvious reasons, such an experiment was not a feasible option. As a result, Milgram was forced to embark

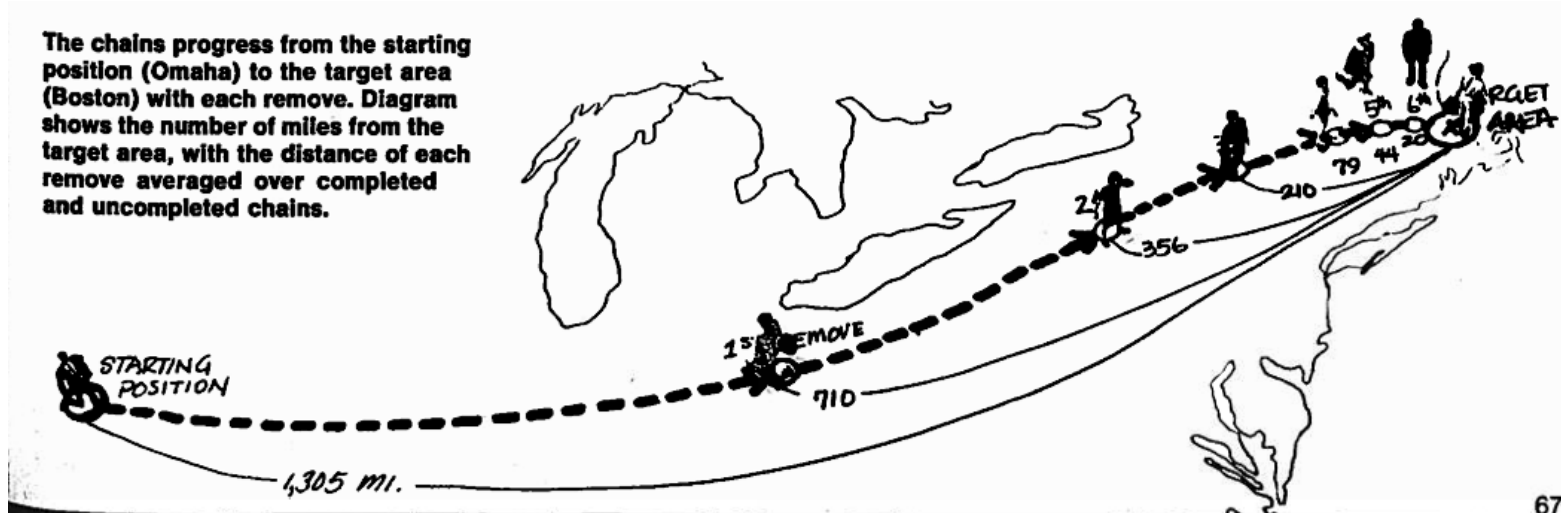


Figure 20.4: An image from Milgram’s original article in *Psychology Today*, showing a “composite” of the successful paths converging on the target person. Each intermediate step is positioned at the average distance of all chains that completed that number of steps. (Image from [297].)

on the much more interesting experiment of constructing paths by “tunneling” through the network, with the letter advancing just one person at a time — a process that could well have failed to reach the target, even if a short path existed.

So the success of the experiment raises fundamental questions about the power of collective search: even if we posit that the social network contains short paths, why should it have been structured so as to make this type of *decentralized search* so effective? Clearly the network contained some type of “gradient” that helped participants guide messages toward the target. As with the Watts-Strogatz model, which sought to provide a simple framework for thinking about short paths in highly clustered networks, this type of search is also something we can try to model: can we construct a random network in which decentralized routing succeeds, and if so, what are the qualitative properties that are crucial for success?

**A model for decentralized search.** To begin with, it is not difficult to model the kind of decentralized search that was taking place in the Milgram experiment. Starting with the grid-based model of Watts and Strogatz, we suppose that a starting node  $s$  is given a message that it must forward to a target node  $t$ , passing it along edges of the network. Initially  $s$  only knows the location of  $t$  on the grid, but, crucially, it does not know the random edges out of any node other than itself. Each intermediate node along the path has this partial information as well, and it must choose which of its neighbors to send the message to next. These choices amount to a collective procedure for finding a path from  $s$  to  $t$  — just as the participants in the Milgram experiment collectively constructed paths to the target person.

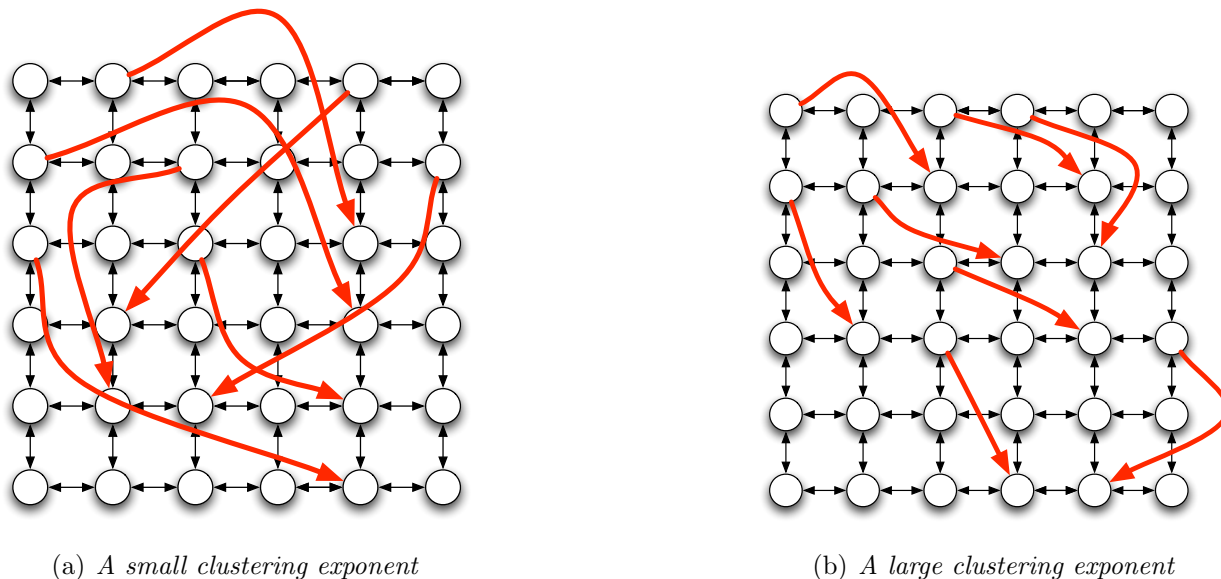


Figure 20.5: With a small clustering exponent, the random edges tend to span long distances on the grid; as the clustering exponent increases, the random edges become shorter.

We will evaluate different search procedures according to their *delivery time* — the expected number of steps required to reach the target, over a randomly generated set of long-range contacts, and randomly chosen starting and target nodes.

Unfortunately, given this set-up, one can prove that decentralized search in the Watts-Strogatz model will necessarily require a large number of steps to reach a target — much larger than the true length of the shortest path [248]. As a mathematical model, the Watts-Strogatz network is thus effective at capturing the density of triangles and the existence of short paths, but not the ability of people, working together in the network, to actually find the paths. Essentially, the problem is that the weak ties that make the world small are “too random” in this model: since they’re completely unrelated to the similarity among nodes that produces the homophily-based links, they’re hard for people to use reliably.

One way to think about this is in terms of Figure 20.4, a hand-drawn image from Milgram’s original article in *Psychology Today*. In order to reach a far-away target, one must use long-range weak ties in a fairly structured, methodical way, constantly reducing the distance to the target. As Milgram observed in the discussion accompanying this picture, “The geographic movement of the [letter] from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain” [297]. So it is not enough to have a network model in which weak ties span only the very long ranges; it is necessary to span all the intermediate ranges of scale as well. Is there a simple way to adapt the model to take this into account?



## 20.4 Modeling the Process of Decentralized Search

Although the Watts-Strogatz model does not provide a structure where decentralized search can be performed effectively, a mild generalization of the model in fact exhibits both properties we want: the networks contain short paths, and these short paths can be found using decentralized search [248].

**Generalizing the network model.** We adapt the model by introducing one extra quantity that controls the “scales” spanned by the long-range weak ties. We have nodes on a grid as before, and each node still has edges to each other node within  $r$  grid steps. But now, each of its  $k$  random edges is generated in a way that decays with distance, controlled by a *clustering exponent*  $q$  as follows. For two nodes  $v$  and  $w$ , let  $d(v, w)$  denote the number of grid steps between them. (This is their distance if one had to walk along adjacent nodes on the grid.) In generating a random edge out of  $v$ , we have this edge link to  $w$  with probability proportional to  $d(v, w)^{-q}$ .

So we in fact have a different model for each value of  $q$ . The original grid-based model corresponds to  $q = 0$ , since then the links are chosen uniformly at random; and varying  $q$  is like turning a knob that controls how uniform the random links are. In particular, when  $q$  is very small, the long-range links are “too random,” and can’t be used effectively for decentralized search (as we saw specifically for the case  $q = 0$  above); when  $q$  is large, the long-range links are “not random enough,” since they simply don’t provide enough of the long-distance jumps that are needed to create a small world. Pictorially, this variation in  $q$  can be seen in the difference between the two networks in Figure 20.5. Is there an optimal operating point for the network, where the distribution of long-range links is sufficiently balanced between these extremes to allow for rapid decentralized search?

In fact there is. The main result for this model is that, in the limit of large network size, decentralized search is most efficient when  $q = 2$  (so that random links follow an inverse-square distribution). Figure 20.6 shows the performance of a basic decentralized search method across different values of  $q$ , for a network of several hundred million nodes. In keeping with the nature of the result — which only holds in the limit as the network size goes to infinity — decentralized search has about the same efficiency on networks of this size across all exponents  $q$  between 1.5 and 2.0. (And at this size, it’s best for a value of  $q$  slightly below 2.) But the overall trend is already clear, and as the network size increases, the best performance occurs at exponents  $q$  closer and closer to 2.

**A Rough Calculation Motivating the Inverse-Square Network.** It is natural to wonder what’s special about the exponent  $q = 2$  that makes it best for decentralized search. In Section 20.7 at the end of this chapter, we describe a proof that decentralized search is efficient when  $q = 2$ , and sketch why search is more efficient with  $q = 2$  — in the limit of

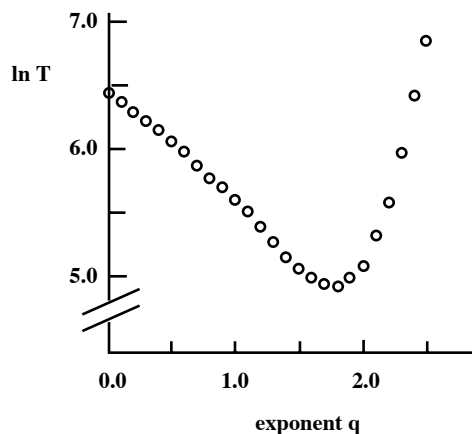


Figure 20.6: Simulation of decentralized search in the grid-based model with clustering exponent  $q$ . Each point is the average of 1000 runs on (a slight variant of) a grid with 400 million nodes. The delivery time is best in the vicinity of exponent  $q = 2$ , as expected; but even with this number of nodes, the delivery time is comparable over the range between 1.5 and 2 [248].

large network size — than with any other exponent. But even without the full details of the proof, there’s a short calculation that suggests why the number 2 is important. We describe this now.

In the real world where the Milgram experiment was conducted, we mentally organize distances into different “scales of resolution”: something can be around the world, across the country, across the state, across town, or down the block. A reasonable way to think about these scales of resolution in a network model — from the perspective of a particular node  $v$  — is to consider the groups of all nodes at increasingly large ranges of distance from  $v$ : nodes at distance 2-4, 4-8, 8-16, and so forth. The connection of this organizational scheme to decentralized search is suggested by Figure 20.4: effective decentralized search “funnels inward” through these different scales of resolution, as we see from the way the letter depicted in this figure reduces its distance to the target by approximately a factor of two with each step.

So now let’s look at how the inverse-square exponent  $q = 2$  interacts with these scales of resolution. We can work concretely with a single scale by taking a node  $v$  in the network, and a fixed distance  $d$ , and considering the group of nodes lying at distances between  $d$  and  $2d$  from  $v$ , as shown in Figure 20.7.

Now, what is the probability that  $v$  forms a link to some node inside this group? Since area in the plane grows like the square of the radius, the total *number* of nodes in this group is proportional to  $d^2$ . On the other hand, the probability that  $v$  links to any one node in the group varies depending on exactly how far out it is, but each individual probability is proportional to  $d^{-2}$ . These two terms — the number of nodes in the group, and the

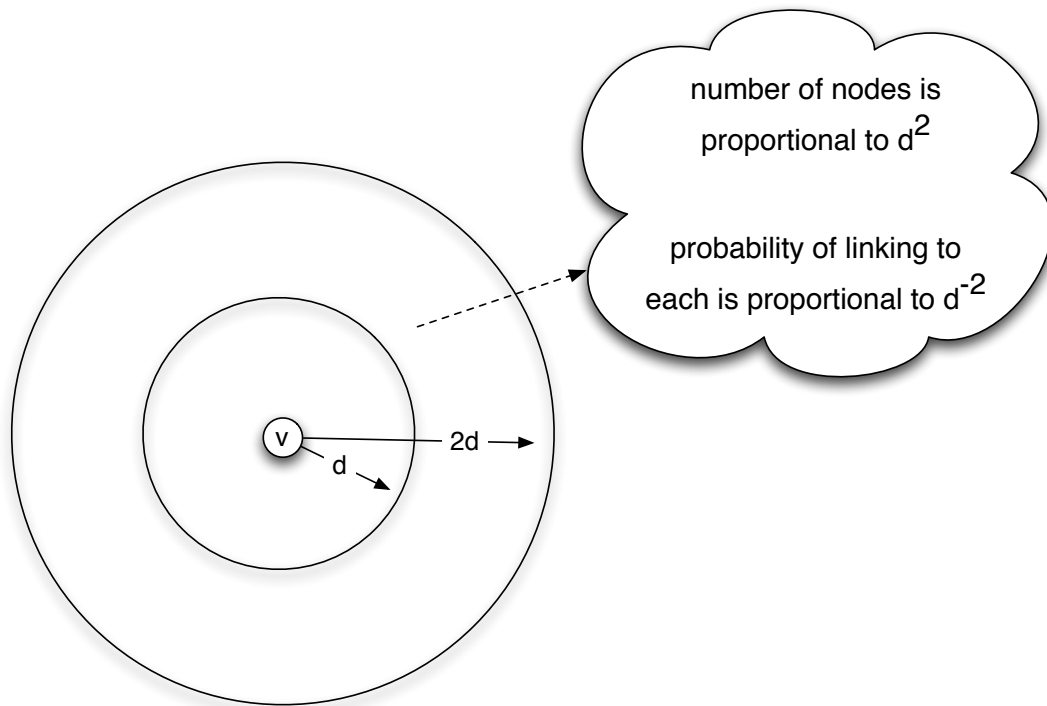


Figure 20.7: The concentric scales of resolution around a particular node.

probability of linking to any one of them — approximately cancel out, and we conclude: the probability that a random edge links into *some node* in this ring is approximately independent of the value of  $d$ .

This, then, suggests a qualitative way of thinking about the network that arises when  $q = 2$ : long-range weak ties are being formed in a way that's spread roughly uniformly over all different scales of resolution. This allows people forwarding the message to consistently find ways of reducing their distance to the target, no matter how near or far they are from it. In this way, it's not unlike how the U.S. Postal Service uses the address on an envelope for delivering a message: a typical postal address exactly specifies scales of resolution, including the country, state, city, street, and finally the street number. But the point is that the postal system is centrally designed and maintained at considerable cost to do precisely this job; the corresponding patterns that guide messages through the inverse-square network are arising spontaneously from a completely random pattern of links.



Figure 20.8: The population density of the LiveJournal network studied by Liben-Nowell et al. (Image from [277].)

## 20.5 Empirical Analysis and Generalized Models

The results we’ve seen thus far have been for stylized models, but they raise a number of qualitative issues that one can try corroborating with data from real social networks. In this section we discuss empirical studies that analyze geographic data to look for evidence of the exponent  $q = 2$ , as well as more general versions of these models that incorporate non-geographic notions of social distance.

**Geographic Data on Friendship.** In the past few years, the rich data available on social networking sites has made it much easier to get large-scale data that provides insight into how friendship links scale with distance. Liben-Nowell et al. [277] used the blogging site LiveJournal for precisely this purpose, analyzing roughly 500,000 users who provided a U.S. ZIP code for their home address, as well as links to their friends on the system. Note that LiveJournal is serving here primarily as a very useful “model system,” containing data on the geographic basis of friendship links on a scale that would be enormously difficult to obtain by more traditional survey methods. From a methodological point of view, it is an interesting and fairly unresolved issue to understand how closely the structure of friendships defined in on-line communities corresponds to the structure of friendships as we understand them in off-line settings.

A number of things have to be done in order to align the LiveJournal data with the basic grid model, and perhaps the most subtle involves the fact that the population density of the users is extremely non-uniform (as it is for the U.S. as a whole). See Figure 20.8 for a visualization of the population density in the LiveJournal data. In particular, the

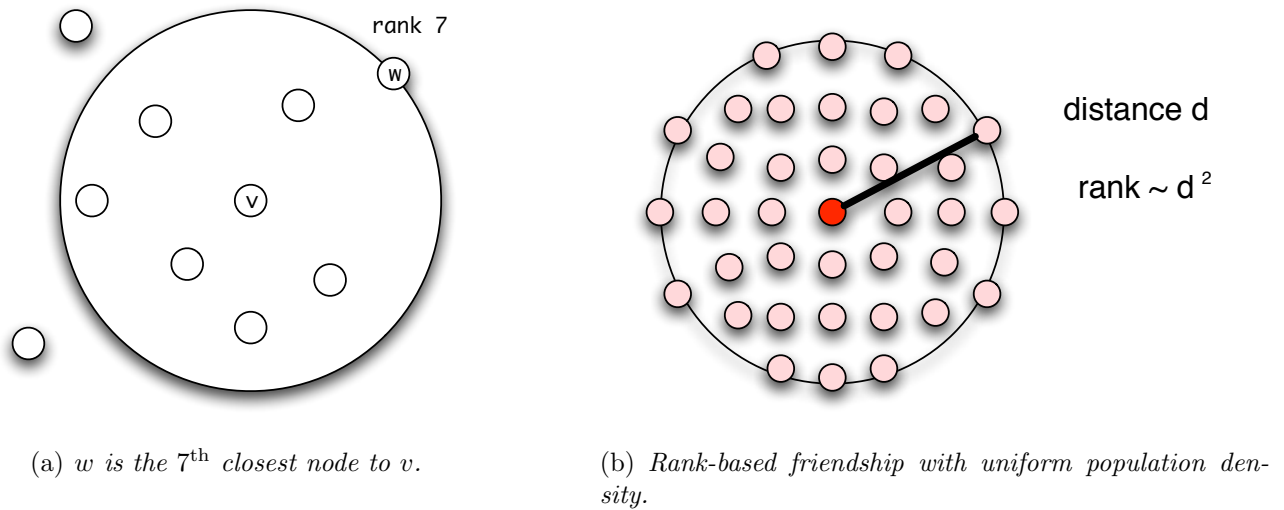


Figure 20.9: When the population density is non-uniform, it can be useful to understand how far  $w$  is from  $v$  in terms of its *rank* rather than its physical distance. In (a), we say that  $w$  has rank 7 with respect to  $v$  because it is the 7<sup>th</sup> closest node to  $v$ , counting outward in order of distance. In (b), we see that for the original case in which the nodes have a uniform population density, a node  $w$  at distance  $d$  from  $v$  will have a rank that is proportional to  $d^2$ , since all the nodes inside the circle of radius  $d$  will be closer to  $v$  than  $w$  is.

inverse-square distribution is useful for finding targets when nodes are uniformly spaced in two dimensions; what's a reasonable generalization to the case in which they can be spread very non-uniformly?

**Rank-Based Friendship.** One approach that works well is to determine link probabilities not by physical distance, but by *rank*. Let's suppose that as a node  $v$  looks out at all other nodes, it *ranks* them by proximity: the rank of a node  $w$ , denoted  $\text{rank}(w)$ , is equal to the number of other nodes that are closer to  $v$  than  $w$  is. For example, in Figure 20.9(a), node  $w$  would have rank seven, since seven other nodes (including  $v$  itself) are closer to  $v$  than  $w$  is. Now, suppose that for some exponent  $p$ , node  $v$  creates a random link as follows: it chooses a node  $w$  as the other end with probability proportional to  $\text{rank}(w)^{-p}$ . We will call this *rank-based friendship* with exponent  $p$ .

Which choice of exponent  $p$  would generalize the inverse-square distribution for uniformly-spaced nodes? As Figure 20.9(b) shows, if a node  $w$  in a uniformly-spaced grid is at distance  $d$  from  $v$ , then it lies on the circumference of a disc of radius  $d$ , which contains about  $d^2$  closer nodes — so its rank is approximately  $d^2$ . Thus, linking to  $w$  with probability proportional to  $d^{-2}$  is approximately the same as linking with probability  $\text{rank}(w)^{-1}$ , so this suggests that exponent  $p = 1$  is the right generalization of the inverse-square distribution. In fact, Liben-Nowell et al. were able to prove that for essentially any population density, if random

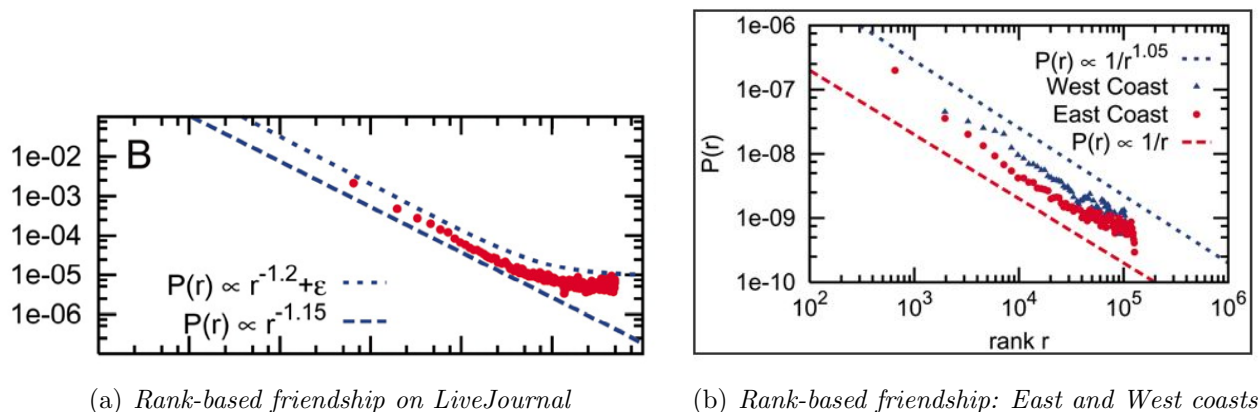


Figure 20.10: The probability of a friendship as a function of geographic rank on the blogging site LiveJournal. (Image from [277].)

links are constructed using rank-based friendship with exponent 1, the resulting network allows for efficient decentralized search with high probability. In addition to generalizing the inverse-square result for the grid, this result has a nice qualitative summary: to construct a network that is efficiently searchable, create a link to each node with probability that is inversely proportional to the number of closer nodes.

Now one can go back to LiveJournal and see how well rank-based friendship fits the distribution of actual social network links: we consider pairs of nodes where one assigns the other a rank of  $r$ , and we ask what fraction  $f$  of these pairs are actually friends, as a function of  $r$ . Does this fraction decrease approximately like  $r^{-1}$ ? Since we're looking for a power-law relationship between the rank  $r$  and the fraction of edges  $f$ , we can proceed as in Chapter 18: rather than plotting  $f$  as a function of  $r$ , we can plot  $\log f$  as a function of  $\log r$ , see if we find an approximately straight line, and then estimate the exponent  $p$  as the slope of this line.

Figure 20.10(a) shows this result for the LiveJournal data; we see that much of the body of the curve is approximately a straight line sandwiched between slopes of  $-1.15$  and  $-1.2$ , and hence close to the optimal exponent of  $-1$ . It is also interesting to work separately with the more structurally homogeneous subsets of the data consisting of West-Coast users and East-Coast users, and when one does this the exponent becomes very close to the optimal value of  $-1$ . Figure 20.10(b) shows this result: The lower dotted line is what you should see if the points followed the distribution  $\text{rank}^{-1}$ , and the upper dotted line is what you should see if the points followed the distribution  $\text{rank}^{-1.05}$ . The proximity of the rank-based exponent on real networks to the optimal value of  $-1$  has also been corroborated by subsequent research. In particular, as part of a recent large-scale study of several geographic phenomena in the Facebook social network, Backstrom et al. [33] returned to the question of rank-based friendship and again found an exponent very close to  $-1$ ; in their case, the

bulk of the distribution was closely approximated by  $\text{rank}^{-0.95}$ .

The plots in Figure 20.10, and their follow-ups, are thus the conclusion of a sequence of steps in which we start from an experiment (Milgram's), build mathematical models based on this experiment (combining local and long-range links), make a prediction based on the models (the value of the exponent controlling the long-range links), and then validate this prediction on real data (from LiveJournal and Facebook, after generalizing the model to use rank-based friendship). This is very much how one would hope for such an interplay of experiments, theories, and measurements to play out. But it is also a bit striking to see the close alignment of theory and measurement in this particular case, since the predictions come from a highly simplified model of the underlying social network, yet these predictions are approximately borne out on data arising from real social networks.

Indeed, there remains a mystery at the heart of these findings. While the fact that the distributions are so close does not necessarily imply the existence of any particular organizing mechanism [70], it is still natural to ask why real social networks have arranged themselves in a pattern of friendships across distance that is close to optimal for forwarding messages to far-away targets. Furthermore, whatever the users of LiveJournal and Facebook are doing, they are not explicitly trying to run versions of the Milgram experiment — if there are dynamic forces or selective pressures driving the network toward this shape, they must be more implicit, and it remains a fascinating open problem to determine whether such forces exist and how they might operate. One intriguing approach to this question has been suggested by Oskar Sandberg, who analyzes a model in which a network constantly re-wires itself as people perform decentralized searches in it. He argues that over time, the network essentially begins to “adapt” to the pattern of searches; eventually the searches become more efficient, and the arrangement of the long-range links begins to approach a structure that can be approximated by rank-based friendship with the optimal exponent [361].

**Social Foci and Social Distance.** When we first discussed the Watts-Strogatz model in Section 20.2, we noted that the grid of nodes was intended to serve as a stylized notion of similarity among individuals. Clearly it is most easily identified with geographic proximity, but subsequent models have explored other types of similarity and the ways in which they can produce small-world effects in networks [250, 410].

The notion of *social foci* from Chapter 4 provides a flexible and general way to produce models of networks exhibiting both an abundance of short paths and efficient decentralized search. Recall that a social focus is any type of community, occupational pursuit, neighborhood, shared interest, or activity that serves to organize social life around it [161]. Foci are a way of summarizing the many possible reasons that two people can know each other or become friends: because they live on the same block, work at the same company, frequent the same cafe, or attend the same kinds of concerts. Now, two people may have many possi-

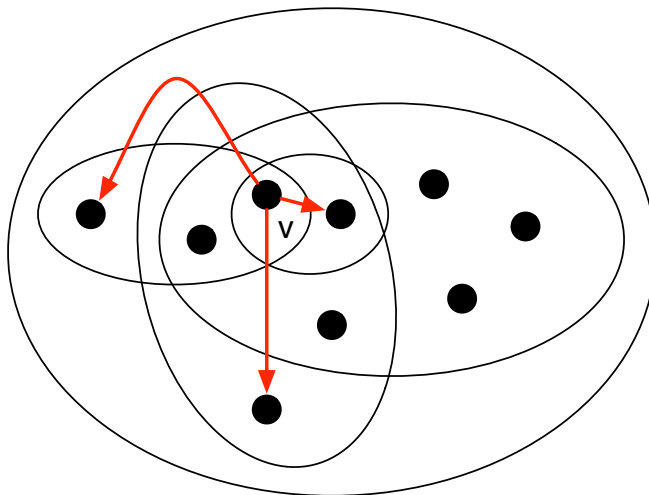


Figure 20.11: When nodes belong to multiple foci, we can define the social distance between two nodes to be the smallest focus that contains both of them. In the figure, the foci are represented by ovals; the node labeled  $v$  belongs to five foci of sizes 2, 3, 5, 7, and 9 (with the largest focus containing all the nodes shown).

ble foci in common, but all else being equal, it is likely that the shared foci with only a few members are the strongest generators of new social ties. For example, two people may both work for the same thousand-person company and live in the same million-person city, but it is the fact that they both belong to the same twenty-person literacy tutoring organization that makes it most probable they know each other. Thus, a natural way to define the *social distance* between two people is to declare it to be *the size of the smallest focus that includes both of them*.

In the previous sections, we've used models that build links in a social network from an underlying notion of geographic distance. Let's consider how this might work with this more general notion of social distance. Suppose we have a collection of nodes, and a collection of foci they belong to — each focus is simply a set containing some of the nodes. We let  $\text{dist}(v, w)$  denote the social distance between nodes  $v$  and  $w$  as defined in terms of shared foci:  $\text{dist}(v, w)$  is the size of the smallest focus that contains both  $v$  and  $w$ . Now, following the style of earlier models, let's construct a link between each pair of nodes  $v$  and  $w$  with probability proportional to  $\text{dist}(v, w)^{-p}$ . For example, in Figure 20.11, the node labeled  $v$  construct links to three other nodes at social distances 2, 3, and 5. One can now show, subject to some technical assumptions on the structure of the foci, that when links are generated this way with exponent  $p = 1$ , the resulting network supports efficient decentralized search with high probability [250].

There are aspects of this result that are similar to what we've just seen for rank-based



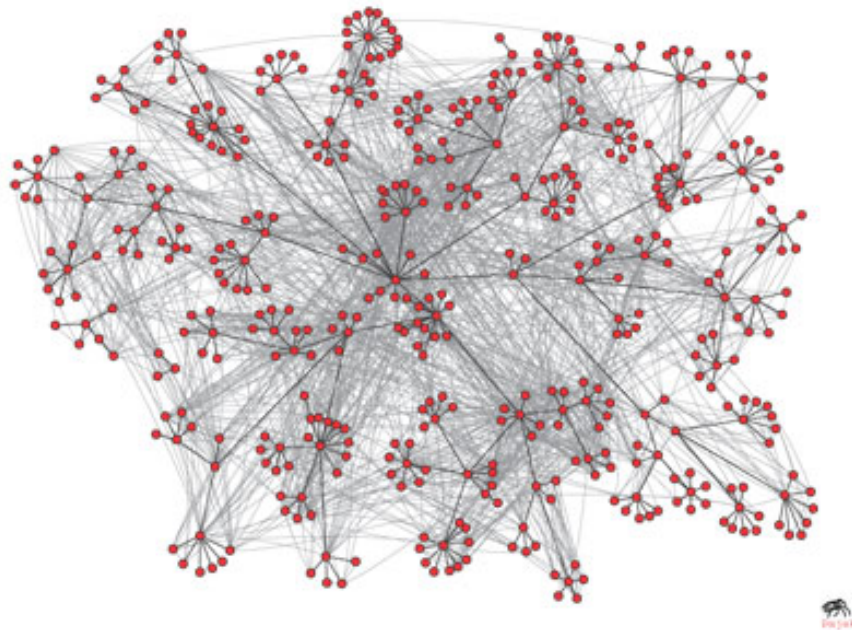


Figure 20.12: The pattern of e-mail communication among 436 employees of Hewlett Packard Research Lab is superimposed on the official organizational hierarchy, showing how network links span different social foci [6]. (Image from <http://www-personal.umich.edu/~ladamic/img/hplabsemailhierarchy.jpg>)

friendship. First, as with rank-based friendship, there is a simple description of the underlying principle: when nodes link to each other with probability inversely proportional to their social distance, the resulting network is efficiently searchable. And second, the exponent  $p = 1$  is again the natural generalization of the inverse-square law for the simple grid model. To see why, suppose we take a grid of nodes and define a set of foci as follows: for each location  $v$  on the grid, and each possible radius  $r$  around that location, there is a focus consisting of all nodes who are within distance  $r$  of  $v$ . (Essentially, these are foci consisting of everyone who live together in neighborhoods and locales of various sizes.) Then for two nodes who are a distance  $d$  apart, their smallest shared focus has a number of nodes proportional to  $d^2$ , so this is their social distance. Thus, linking with probability proportional to  $d^{-2}$  is essentially the same as linking with probability inversely proportional to their social distance.

Recent studies of who-talks-to-whom data has fit this model to social network structures. In particular, Adamic and Adar analyzed a social network on the employees of Hewlett Packard Research Lab that we discussed briefly in Chapter 1 connecting two people if they exchanged e-mail at least six times over a three-month period [6]. (See Figure 20.12.) They then defined a focus for each of the groups within the organizational structure (i.e. a group of employees all reporting a common manager). They found that the probability of a link

between two employees at social distance  $d$  within the organization scaled proportionally to  $d^{-3/4}$ . In other words, the exponent on the probability for this network is close to, but smaller than, the best exponent for making decentralized search within the network efficient.

These increasingly general models thus provide us with a way to look at social-network data and speak quantitatively about the ways in which the links span different levels of distance. This is important not just for understanding the small-world properties of these networks, but also more generally for the ways in which homophily and weak ties combine to produce the kinds of structures we find in real networks.

**Search as an Instance of Decentralized Problem-Solving.** While the Milgram experiment was designed to test the hypothesis that people are connected by short paths in the global social network, our discussion here shows that it also served as an empirical study of people’s ability to collectively solve a problem — in this case, searching for a path to a far-away individual — using only very local information, and by communicating only with their neighbors in the social network. In addition to the kinds of search methods discussed here, based on aiming as closely to the target as possible in each step, researchers have also studied the effectiveness of path-finding strategies in which people send messages to friends who have a particularly large number of edges (on the premise that they will be “better connected” in general) [6, 7], as well as strategies that explicitly trade off the proximity of a person against their number of edges [370].

The notion that social networks can be effective at this type of decentralized problem-solving is an intriguing and general premise that applies more broadly than just to the problem of path-finding that Milgram considered. There are many possible problems that people interacting in a network could try solving, and it is natural to suppose that their effectiveness will depend both on the difficulty of the problem being solved and on the network that connects them. There is a long history of experimental interest in collective problem-solving [47], and indeed one way to view the bargaining experiments described in Chapter 12 is as an investigation of the ability of a group of people to collectively find a mutually compatible set of exchanges when their interaction is constrained by a network. Recent experiments have explored this issue for a range of basic problems, across multiple kinds of network structures [236, 237], and there is also a growing line of work in the design of systems that can exploit the power of collective human problem-solving by very large on-line populations [402, 403].

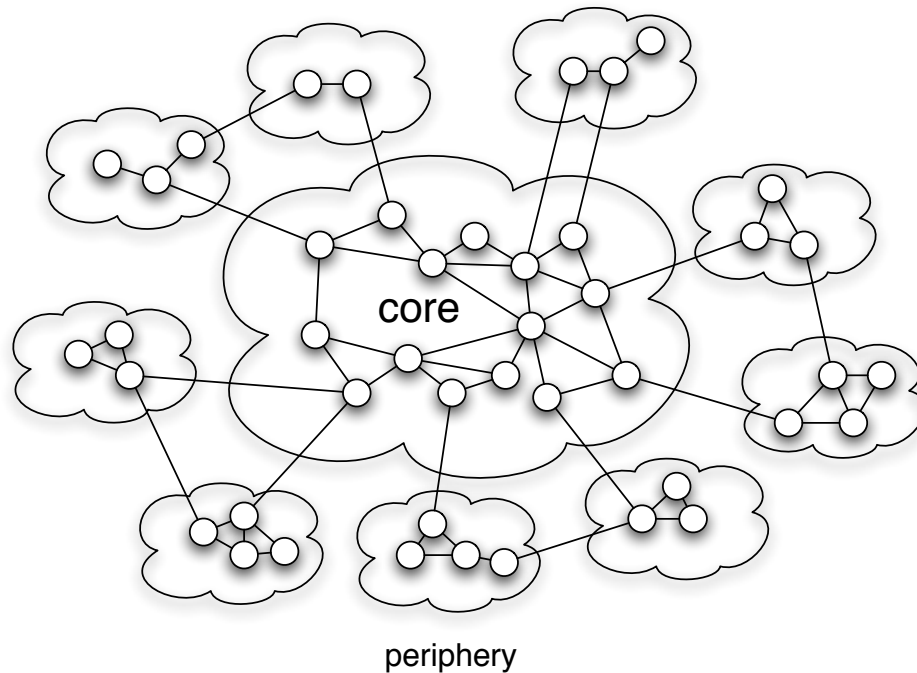


Figure 20.13: The core-periphery structure of social networks.

## 20.6 Core-Periphery Structures and Difficulties in Decentralized Search

In the four decades since the Milgram experiment, the research community has come to appreciate both the robustness and the delicacy of the “six degrees” principle. As we noted in Chapter 2, many studies of large-scale social network data have confirmed the pervasiveness of very short paths in almost every setting. On the other hand, the ability of people to find these paths from within the network is a subtle phenomenon: it is striking that it should happen at all, and the conditions that facilitate it are not fully understood.

As Judith Kleinfeld has noted in her recent critique of the Milgram experiment [255], the success rate at finding targets in recreations of the experiment has often been much lower than it was in the original work. Much of the difficulty can be explained by lack of participation: many people, asked to forward a letter as part of the experiment, will simply throw it away. This is consistent with lack of participation in any type of survey or activity carried out by mail; assuming this process is more or less random, it has a predictable effect on the results, and one can correct for it [131, 416].

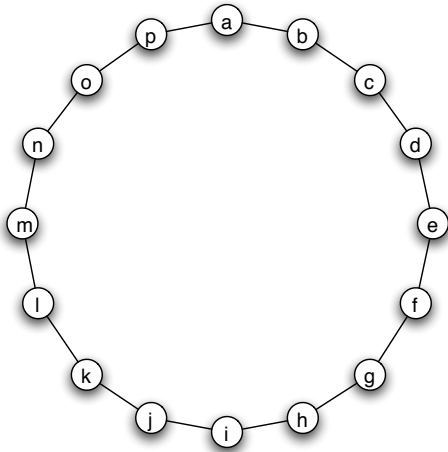
But there are also more fundamental difficulties at work, pointing to questions about large social networks that may help inform a richer understanding of network structure. In

particular, Milgram-style search in a network is most successful when the target person is affluent and socially high-status. For example, in the largest small-world experiment to date [131], 18 different targets were used, drawn from a wide range of backgrounds. Completion rates to all targets were small, due to lack of participation in the e-mail based forwarding of messages, but they were highest for targets who were college professors and journalists, and particularly small for low-status targets.

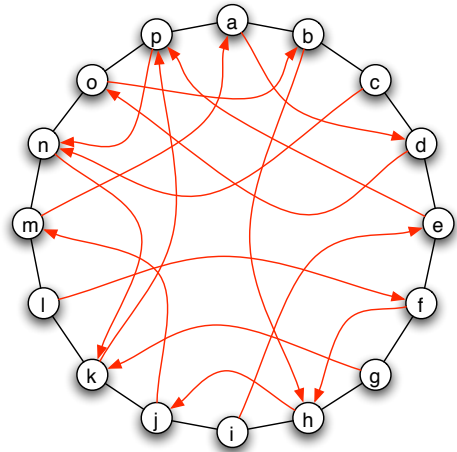
**Core-Periphery Structures.** This wide variation in the success rates of search to different targets does not simply arise from variations in individual attributes of the respective people — it is based on the fact that social networks are structured to make high-status individuals much easier to find than low-status ones. Homophily suggests that high-status people will mainly know other high-status people, and low-status people will mainly know other low-status people, but this does not imply that the two groups occupy symmetric or interchangeable positions in the social network. Rather, large social networks tend to be organized in what is called a *core-periphery structure* [72], in which the high-status people are linked in a densely-connected *core*, while the low-status people are atomized around the *periphery* of the network. Figure 20.13 gives a schematic picture of such a structure. High-status people have the resources to travel widely; to meet each other through shared foci around clubs, interests, and educational and occupational pursuits; and more generally to establish links in the network that span geographic and social boundaries. Low-status people tend to form links that are much more clustered and local. As a result, the shortest paths connecting two low-status people who are geographically or socially far apart will tend to go into the core and then come back out again.

All this has clear implications for people's ability to find paths to targets in the network. In particular, it indicates some of the deep structural reasons why it is harder for Milgram-style decentralized search to find low-status targets than high-status targets. As you move toward a high-status target, the link structure tends to become richer, based on connections with an increasing array of underlying social reasons. In trying to find a low-status target, on the other hand, the link structure becomes structurally more impoverished as you move toward the periphery.

These considerations suggest an opportunity for richer models that take status effects more directly into account. The models we have seen capture the process by which people can find each other when they are all embedded in an underlying social structure, and motivated to continue a path toward a specific destination. But as the social structure begins to fray around the periphery, an understanding of how we find our way through it has the potential to shed light not just on the networks themselves, but on the way that network structure is intertwined with status and the varied positions that different groups occupy in society as a whole.



(a) A set of nodes arranged in a ring.



(b) A ring augmented with random long-range links.

Figure 20.14: The analysis of decentralized search is a bit cleaner in one dimension than in two, although it is conceptually easy to adapt the arguments to two dimensions. As a result, we focus most of the discussion on a one-dimensional ring augmented with random long-range links.

## 20.7 Advanced Material: Analysis of Decentralized Search

In Section 20.4, we gave some basic intuition for why an inverse-square distribution of links with distance makes effective decentralized search possible. Even given this way of thinking about it, however, it still requires further work to really see why search succeeds with this distribution. In this section, we describe the complete analysis of the process [249].

To make the calculations a bit simpler, we vary the model in one respect: we place the nodes in one dimension rather than two. In fact, the argument is essentially the same no matter how many dimensions the nodes are in, but one dimension makes things the cleanest (even if not the best match for the actual geographic structure of a real population). It turns out, as we will argue more generally later in this section, that the best exponent for search is equal to the dimension, so in our one-dimensional analysis we will be using an exponent of  $q = 1$  rather than  $q = 2$ . At the end, we will discuss the minor ways in which the argument needs to be adapted in two or higher dimensions.

We should also mention, recalling the discussion earlier in the chapter, that there is a second fundamental part of this analysis as well — showing that this choice of  $q$  is in fact the best for decentralized search in the limit of increasing network size. At the end, we sketch why this is true, but the full details are beyond what we will cover here.

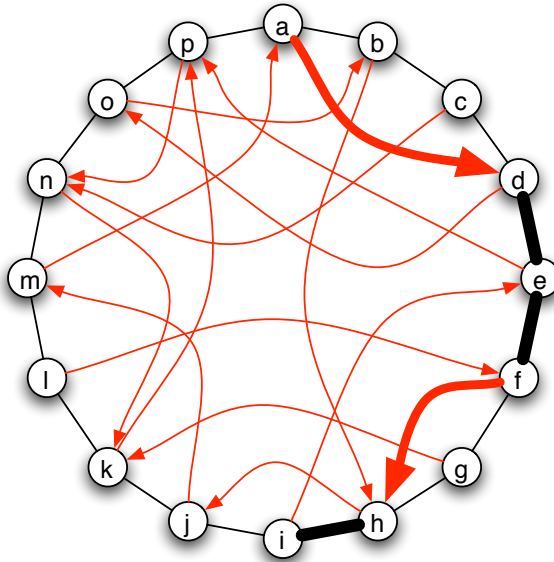


Figure 20.15: In myopic search, the current message-holder chooses the contact that lies closest to the target (as measured on the ring), and it forwards the message to this contact.

## A. The Optimal Exponent in One Dimension

Here, then, is the model we will be looking at. A set of  $n$  nodes are arranged on a one-dimensional ring as shown in Figure 20.14(a), with each node connected by directed edges to the two others immediately adjacent to it. Each node  $v$  also has a single directed edge to some other node on the ring; the probability that  $v$  links to any particular node  $w$  is proportional to  $d(v, w)^{-1}$ , where  $d(v, w)$  is their distance apart on the ring. We will call the nodes to which  $v$  has an edge its *contacts*: the two nodes adjacent to it on the ring are its *local contacts*, and the other one is its *long-range contact*. The overall structure is thus a ring that is augmented with random edges, as shown in Figure 20.14(b). Again, this is essentially just a one-dimensional version of the grid with random edges that we saw in Figure 20.5.<sup>1</sup>

**Myopic Search.** Let's choose a random start node  $s$  and a random target node  $t$  on this augmented ring network. The goal, as in the Milgram experiment, is to forward a message from the start to the target, with each intermediate node on the way only knowing the locations of its own neighbors, and the location of  $t$ , but nothing else about the full network.

The forwarding strategy that we analyze, which works well on the ring when  $q = 1$ , is a

<sup>1</sup>We could also analyze a model in which nodes have more outgoing edges, but this only makes the search problem easier; our result here will show that even when each node has only two local contacts and a single long-range contact, search can still be very efficient.

simple technique that we call *myopic search*: when a node  $v$  is holding the message, it passes it to the contact that lies as close to  $t$  on the ring as possible. Myopic search can clearly be performed even by nodes that know nothing about the network other than the locations of their friends and the location of  $t$ , and it is a reasonable approximation to the strategies used by most people in Milgram-style experiments [243].

For example, Figure 20.15 shows the myopic path that would be constructed if we chose  $a$  as the start node and  $i$  as the target node in the network from Figure 20.14(b).

1. Node  $a$  first sends the message to node  $d$ , since among  $a$ 's contacts  $p$ ,  $b$ , and  $d$ , node  $d$  lies closest to  $i$  on the ring.
2. Then  $d$  passes the message to its local contact  $e$ , and  $e$  likewise passes the message to its local contact  $f$ , since the long-range contacts of both  $d$  and  $e$  lead away from  $i$  on the ring, not closer to it.
3. Node  $f$  has a long-range contact  $h$  that proves useful, so it passes it to  $h$ . Node  $h$  actually has the target as a local contact, so it hands it directly to  $i$ , completing the path in five steps.

Notice that this myopic path is not the shortest path from  $a$  to  $i$ . If  $a$  had known that its friend  $b$  in fact had  $h$  as a contact, it could have handed the message to  $b$ , thereby taking the first step in the three-step  $a$ - $b$ - $h$ - $i$  path. It is precisely this lack of knowledge about the full network structure that prevents myopic search from finding the true shortest path in general.

Despite this, however, we will see next that in expectation, myopic search finds paths that are surprisingly short.

**Analyzing Myopic Search: The Basic Plan.** We now have a completely well-defined probabilistic question to analyze, as follows. We generate a random network by adding long-range edges to a ring as above. We then choose a random start node  $s$  and random target node  $t$  in this network. The number of steps required by myopic search is now a random variable  $X$ , and we are interested in showing that  $E[X]$ , the expected value of  $X$ , is relatively small.

Our plan for putting a bound on the expected value of  $X$  follows the idea contained in Milgram's picture from Figure 20.4: we track how long it takes for the message to reduce its distance by factors of two as it closes in on the target. Specifically, as the message moves from  $s$  to  $t$ , we'll say that it's in *phase*  $j$  of the search if its distance from the target is between  $2^j$  and  $2^{j+1}$ . See Figure 20.16 for an illustration of this. Notice that the number of different phases is at most  $\log_2 n$  — that is, the number of doublings needed to go from 1 to  $n$ . (In what follows, we will drop the base of the logarithm and simply write  $\log n$  to denote  $\log_2 n$ .)

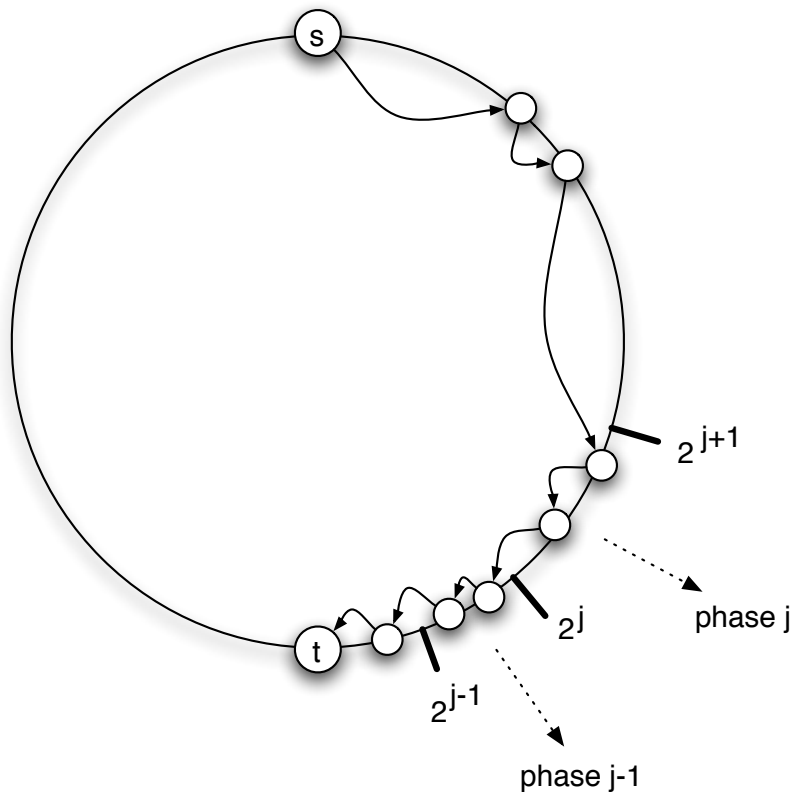


Figure 20.16: We analyze the progress of myopic search in *phases*. Phase  $j$  consists of the portion of the search in which the message's distance from the target is between  $2^j$  and  $2^{j+1}$ .

We can write  $X$ , the number of steps taken by the full search, as

$$X = X_1 + X_2 + \cdots + X_{\log n};$$

that is, the total time taken by the search is simply the sum of the times taken in each phase. Linearity of expectation says that the expectation of a sum of random variables is equal to the sum of their individual expectations, and so we have

$$E[X] = E[X_1 + X_2 + \cdots + X_{\log n}] = E[X_1] + E[X_2] + \cdots + E[X_{\log n}].$$

We will now show — and this is the crux of the argument — that the expected value of each  $X_j$  is at most proportional to  $\log n$ . In this way,  $E[X]$  will be a sum of  $\log n$  terms, each at most proportional to  $\log n$ , and so we will have shown that  $E[X]$  is at most proportional to  $(\log n)^2$ .

This will achieve our overall goal of showing that myopic search is very efficient with the given distribution of links: the full network has  $n$  nodes, but myopic search constructs a path that is *exponentially* smaller: proportional to the square of  $\log n$ .



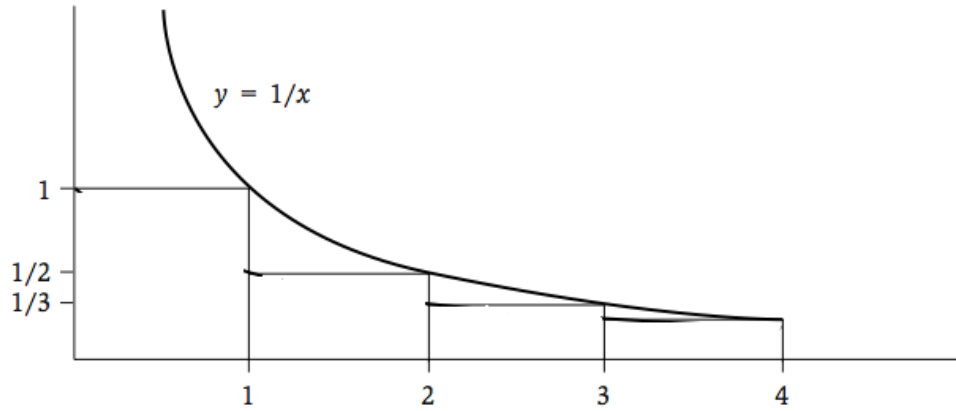


Figure 20.17: Determining the normalizing constant for the probability of links involves evaluating the sum of the first  $n/2$  reciprocals. An upper bound on the value of this sum can be determined from the area under the curve  $y = 1/x$ .

**Intermediate Step: The Normalizing Constant** In implementing this high-level strategy, the first thing we need to work out is in fact something very basic: we've been saying all along that  $v$  forms its long-range link to  $w$  with probability *proportional to*  $d(v, w)^{-1}$ , but what is the constant of proportionality? As in any case when we know a set of probabilities up to a missing constant of proportionality  $1/Z$ , the value of  $Z$  is here simply the sum of  $d(v, u)^{-1}$  over all nodes  $u \neq v$  on the ring. Dividing everything down by this normalizing constant  $Z$ , the probability of  $v$  linking to  $w$  is then equal to  $\frac{1}{Z}d(v, w)^{-1}$ .

To work out the value of  $Z$ , we note that there are two nodes at distance 1 from  $v$ , two at distance 2, and more generally two at each distance  $d$  up to  $n/2$ . Assuming  $n$  is even, there is also a single node at distance  $n/2$  from  $v$  — the node diametrically opposite it on the ring. Therefore, we have

$$Z \leq 2 \left( 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n/2} \right). \quad (20.1)$$

The quantity inside parentheses on the right is a common expression in probabilistic calculations: the sum of the first  $k$  reciprocals, for some  $k$ , in this case  $n/2$ . To put an upper bound on its size, we can compare it to the area under the curve  $y = 1/x$ , as shown in Figure 20.17. As that figure indicates, a sequence of rectangles of unit widths and heights  $1/2, 1/3, 1/4, \dots, 1/k$  fits under the curve  $y = 1/x$  as  $x$  ranges from 1 to  $k$ . Combined with a single rectangle of height and width 1, we see that

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{k} \leq 1 + \int_1^k \frac{1}{x} dx = 1 + \ln k.$$

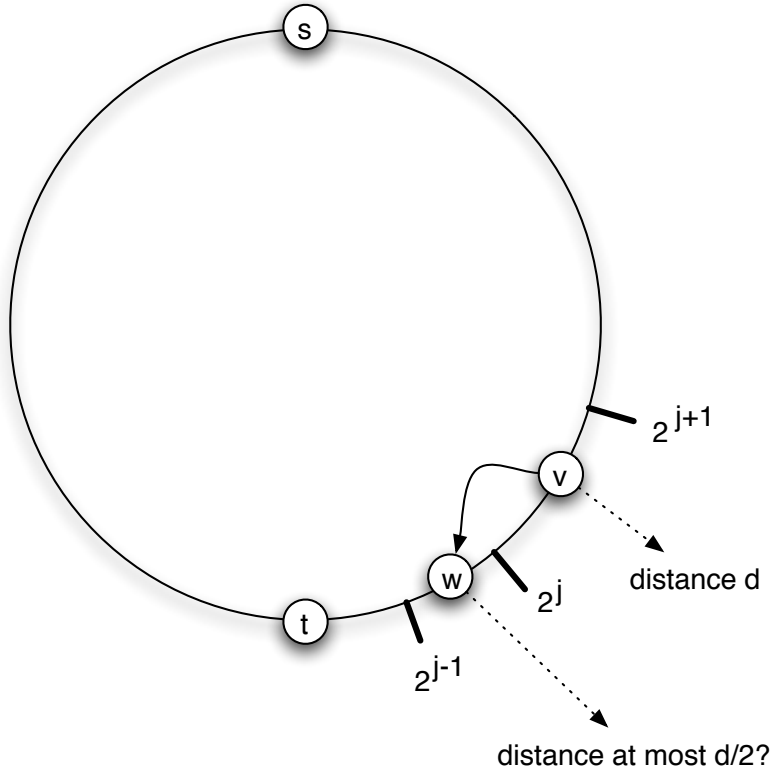


Figure 20.18: At any given point in time, the search is in some phase  $j$ , with the message residing at a node  $v$  at distance  $d$  from the target. The phase will come to an end if  $v$ 's long-range contact lies at distance  $\leq d/2$  from the target  $t$ , and so arguing that the probability of this event is large provides a way to show that the phase will not last too long.

Plugging in  $k = n/2$  to the expression on the right-hand side of inequality (20.1) above, we get

$$Z \leq 2(1 + \ln(n/2)) = 2 + 2\ln(n/2).$$

For simplicity, we'll use a slightly weaker bound on  $Z$ , which follows simply from the observation that  $\ln x \leq \log_2 x$ :

$$Z \leq 2 + 2\log_2(n/2) = 2 + 2(\log_2 n) - 2(\log_2 2) = 2\log_2 n.$$

Thus, we now have an expression for the actual probability that  $v$  links to  $w$  (including its constant of proportionality): the probability  $v$  links to  $w$  is

$$\frac{1}{Z}d(v, w)^{-1} \geq \frac{1}{2\log n}d(v, w)^{-1}.$$

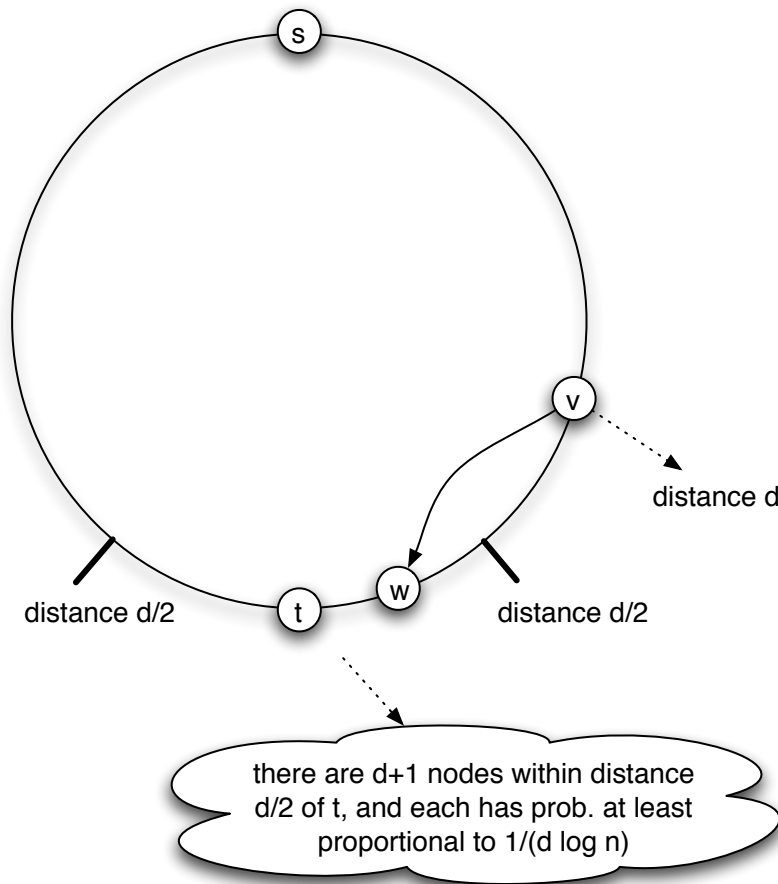


Figure 20.19: Showing that, with reasonable probability,  $v$ 's long-range contact lies within half the distance to the target.

**Analyzing the Time Spent in One Phase of Myopic Search.** Finally, we come to the last and central step of the analysis: showing that the time spent by the search in any one phase is not very large. Let's choose a particular phase  $j$  of the search, when the message is at a node  $v$  whose distance to the target  $t$  is some number  $d$  between  $2^j$  and  $2^{j+1}$ . (See Figure 20.18 for an illustration of all this notation in context.) The phase will come to an end once the distance to the target decreases below  $2^j$ , and we want to show that this will happen relatively quickly.

One way for the phase to come to an end immediately would be for  $v$ 's long-range contact  $w$  to be at distance  $\leq \frac{d}{2}$  from  $t$ . In this case,  $v$  would necessarily be the last node to belong to phase  $j$ . So let's show that this immediate halving of the distance in fact happens with reasonably large probability.

The argument is pictured in Figure 20.19. Let  $I$  be the set of nodes at distance  $\leq \frac{d}{2}$  from

$t$ ; this is where we hope  $v$ 's long-range contact will lie. There are  $d + 1$  nodes in  $I$ : this includes node  $t$  itself, and  $d/2$  nodes consecutively on each side of it. Each node  $w$  in  $I$  has distance at most  $3d/2$  from  $v$ : the farthest one is on the “far side” of  $t$  from  $v$ , at distance  $d + d/2$ . Therefore, each node  $w$  in  $I$  has probability at least

$$\frac{1}{2 \log n} d(v, w)^{-1} \geq \frac{1}{2 \log n} \cdot \frac{1}{3d/2} = \frac{1}{3d \log n}$$

of being the long-range contact of  $v$ . Since there are more than  $d$  nodes in  $I$ , the probability that *one of them* is the long-range contact of  $v$  is at least

$$d \cdot \frac{1}{3d \log n} = \frac{1}{3 \log n}.$$

If one of these nodes is the long-range contact of  $v$ , then phase  $j$  ends immediately in this step. Therefore, in each step that it proceeds, phase  $j$  has a probability of at least  $1/(3 \log n)$  of coming to an end, independently of what has happened so far. To run for at least  $i$  steps, phase  $j$  has to fail to come to an end  $i - 1$  times in a row, and so the probability that phase  $j$  runs for at least  $i$  steps is at most

$$\left(1 - \frac{1}{3 \log n}\right)^{i-1}.$$

Now we conclude by just using the formula for the expected value of a random variable:

$$E[X_j] = 1 \cdot \Pr[X_j = 1] + 2 \cdot \Pr[X_j = 2] + 3 \cdot \Pr[X_j = 3] + \cdots \quad (20.2)$$

There is a useful alternate way to write this: notice that in the expression

$$\Pr[X_j \geq 1] + \Pr[X_j \geq 2] + \Pr[X_j \geq 3] + \cdots \quad (20.3)$$

the quantity  $\Pr[X_j = 1]$  is accounted for once (in the first term only), the quantity  $\Pr[X_j = 2]$  is accounted for twice (in the first two terms only), and so forth. Therefore the expressions in (20.2) and (20.3) are the same thing, and so we have

$$E[X_j] = \Pr[X_j \geq 1] + \Pr[X_j \geq 2] + \Pr[X_j \geq 3] + \cdots \quad (20.4)$$

Now, we've just argued above that

$$\Pr[X_j \geq i] \leq \left(1 - \frac{1}{3 \log n}\right)^{i-1},$$

and so

$$E[X_j] \leq 1 + \left(1 - \frac{1}{3 \log n}\right) + \left(1 - \frac{1}{3 \log n}\right)^2 + \left(1 - \frac{1}{3 \log n}\right)^3 + \cdots$$

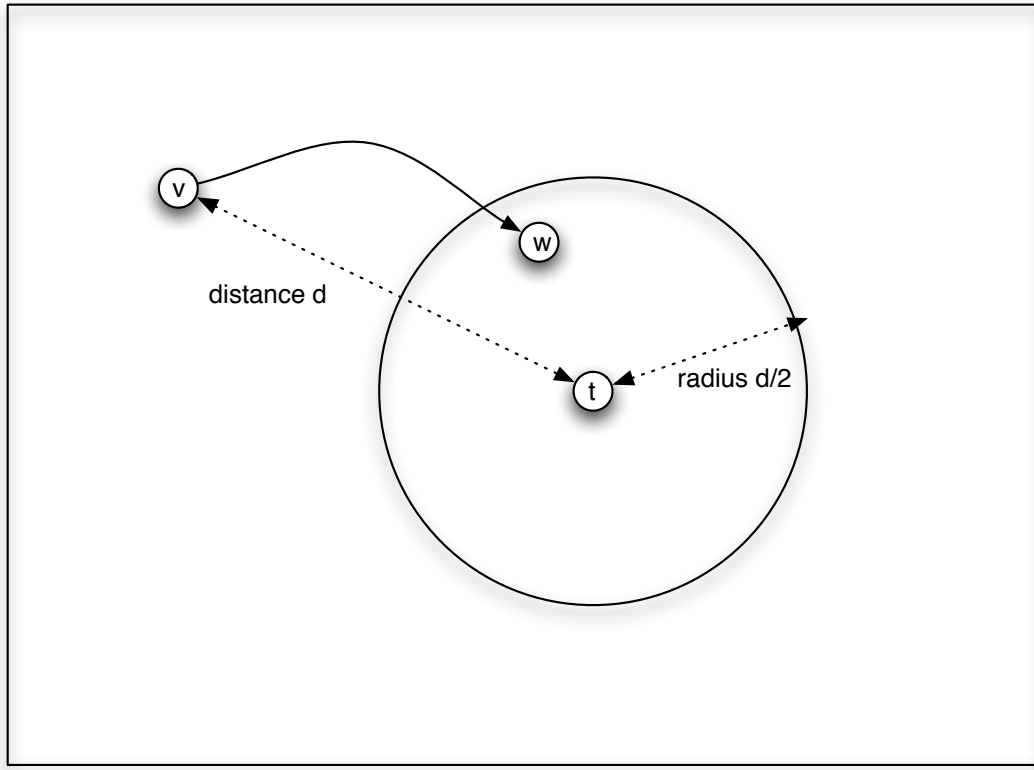


Figure 20.20: The analysis for the one-dimensional ring can be carried over almost directly to the two-dimensional grid. In two dimensions, with the message at a current distance  $d$  from the target  $t$ , we again look at the set of nodes within distance  $d/2$  of  $t$ , and argue that the probability of entering this set in a single step is reasonably large.

The right-hand side is a geometric sum with multiplier  $\left(1 - \frac{1}{3 \log n}\right)$ , and so it converges to

$$\frac{1}{1 - \left(1 - \frac{1}{3 \log n}\right)} = 3 \log n.$$

Thus we have

$$E[X_j] \leq 3 \log n.$$

And now we're done.  $E[X]$  is a sum of the  $\log n$  terms  $E[X_1] + E[X_2] + \cdots + E[X_{\log n}]$ , and we've just argued that each of them is at most  $3 \log n$ . Therefore,  $E[X] \leq 3(\log n)^2$ , a quantity proportional to  $(\log n)^2$  as we wanted to show.

## B. Higher Dimensions and Other Exponents

Using the analysis we've just completed, we now discuss two further issues. First, we sketch how it can be used to analyze networks built by adding long-range contacts to nodes arranged in two dimensions. Then we show how, in the limit of increasing network size, search is more efficient when  $q$  is equal to the underlying dimension than when it is equal to any other value.

**The Analysis in Two Dimensions.** It's not hard to adapt our analysis for the one-dimensional ring directly to the case of the two-dimensional grid. Essentially, we only used the fact that we were in one dimension in two distinct places in the analysis. First, we used it when we determined the normalizing constant  $Z$ . Second, and in the end most crucially, we used it to argue that there were at least  $d$  nodes within distance  $d/2$  of the target  $t$ . This factor of  $d$  canceled the  $d^{-1}$  in the link probability, allowing us to conclude that the probability of halving the distance to the target in any given step was at least proportional to  $1/(\log n)$ , *regardless of the value of  $d$* .

At a qualitative level, this last point is the heart of the analysis: with link probability  $d^{-1}$  on the ring, the probability of linking to any one node exactly offsets the number of nodes close to  $t$ , and so myopic search makes progress at every possible distance away from the target.

When we go to two dimensions, the number of nodes within distance  $d/2$  of the target will be proportional to  $d^2$ . This suggests that to get the same nice cancellation property, we should have  $v$  link to each node  $w$  with probability proportional to  $d(v, w)^{-2}$ , and this exponent  $-2$  is what we will use.

With the above ideas in mind, and with this change in the exponent to  $-2$ , the analysis for two dimensions is almost exactly the same as what we just saw for the one-dimensional ring. First, while we won't go through the calculations here, the normalizing constant  $Z$  is still proportional to  $\log n$  when the probability of  $v$  linking to  $w$  is proportional to  $d(v, w)^{-2}$ . We then consider  $\log n$  different phases as before; and as depicted in Figure 20.20, we consider the probability that at any given moment, the current message-holder  $v$  has a long-range contact  $w$  that halves the distance to the target, ending the phase immediately. Now we use the calculation foreshadowed in the previous paragraph: the number of nodes within distance  $d/2$  of the target is proportional to  $d^2$ , and the probability that  $v$  links to each is proportional to  $1/(d^2 \log n)$ . Therefore, the probability that the message halves its distance to the target in this step is at least  $d^2/(d^2 \log n) = 1/(\log n)$ , and the rest of the analysis then finishes as before.

A similarly direct adaptation of the analysis shows that decentralized search is efficient for networks built by adding long-range contacts to grids in  $D > 2$  dimensions, when the exponent  $q$  is equal to  $D$ .

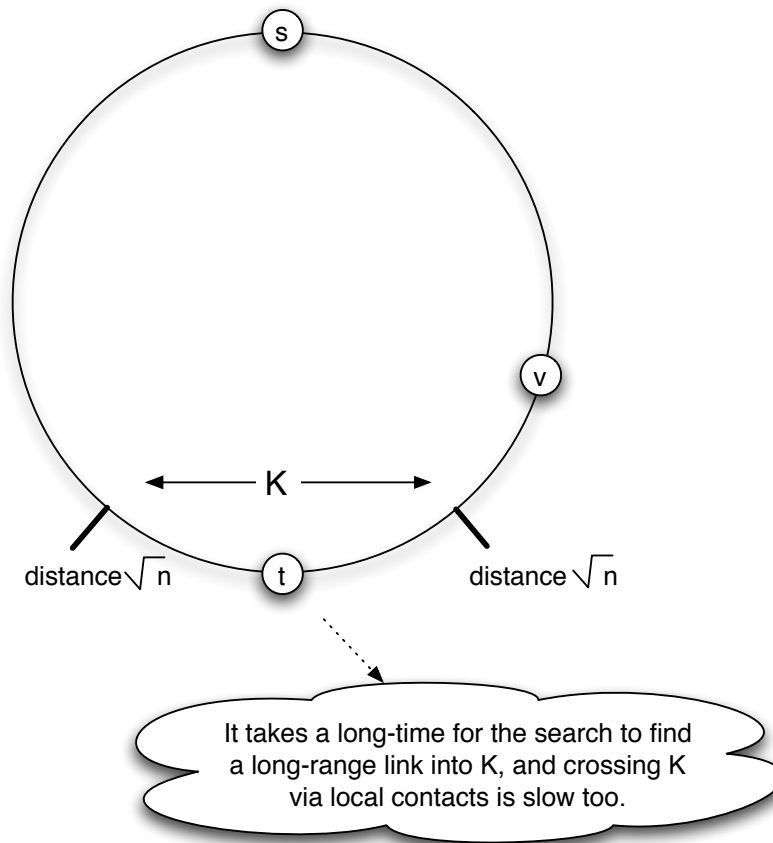


Figure 20.21: To show that decentralized search strategies require large amounts of time with exponent  $q = 0$ , we argue that it is difficult for the search to cross the set of  $\sqrt{n}$  nodes closest to the target. Similar arguments hold for other exponents  $q < 1$ .

**Why Search is Less Efficient with Other Exponents.** Finally, let's sketch why decentralized search is less efficient when the exponent is anything else. For concreteness, we'll focus on why search doesn't work well when  $q = 0$  — the original Watts-Strogatz model when long-range links are chosen uniformly at random. Also, we'll talk again about the one-dimensional ring rather than the two-dimensional grid, since things are a bit cleaner in one dimension, although again the analysis in two dimensions is essentially the same.

The key idea, as with the “good” exponent  $q = 1$ , is to consider the set of all nodes within some distance of the target  $t$ . But whereas in the case of  $q = 1$  we wanted to argue that it is easy to enter smaller and smaller sets centered around  $t$ , here we want to identify a set of nodes centered at  $t$  that is somehow “impenetrable” — a set that is very hard for the search to enter.

In fact, it is not difficult to do this. The basic idea is depicted in Figure 20.21; we sketch

how the argument works, but without going into all the details. (The details can be found in [249].) Let  $K$  be the set of all nodes within distance less than  $\sqrt{n}$  of the target  $t$ . Now, with high probability, the starting point of the search lies outside  $K$ . Because long-range contacts are created uniformly at random (since  $q = 0$ ), the probability that any one node has a long-range contact inside  $K$  is equal to the size of  $K$  divided by  $n$ : so it is less than  $2\sqrt{n}/n = 2/\sqrt{n}$ . Therefore, any decentralized search strategy will need at least  $\sqrt{n}/2$  steps in expectation to find a node with a long-range contact in  $K$ . On the other hand, as long as it doesn't find a long-range link leading into  $K$ , it can't reach the target in less than  $\sqrt{n}$  steps, since it would take this long to “walk” step-by-step through  $K$  using only the connections among local contacts. From this, one can show that the expected time for any decentralized search strategy to reach  $t$  must be at least proportional to  $\sqrt{n}$ .

There are similar arguments for every other exponent  $q \neq 1$ . When  $q$  is strictly between 0 and 1, a version of the argument above works, with a set  $K$  centered at  $t$  whose width depends on the value of  $q$ . And when  $q > 1$ , decentralized search is inefficient for a different reason: since even the long-range links are relatively short, it takes a long time for decentralized search to find links that span sufficiently long distances. This makes it hard to quickly traverse the distance from the starting node to the target.

Overall, one can show that for every exponent  $q \neq 1$ , there is a constant  $c > 0$  (depending on  $q$ ), so that it takes at least proportional to  $n^c$  steps in expectation for any decentralized search strategy to reach the target in a network generated with exponent  $q$ . So in the limit, as  $n$  becomes large, decentralized search with exponent  $q = 1$  requires time that grows like a polynomial in  $\log n$ , while decentralized search at any other exponent requires a time that grows like a polynomial in  $n$  — exponentially worse.<sup>2</sup> The exponent  $q = 1$  on the ring — or  $q = 2$  in the plane — is optimally balanced between producing networks that are “too random” for search, and those that are not random enough.

## 20.8 Exercises

1. In the basic “six degrees of separation” question, one asks whether most pairs of people in the world are connected by a path of at most six edges in the social network, where an edge joins any two people who know each other on a first-name basis.

Now let's consider a variation on this question. Suppose that we consider the full population of the world, and suppose that from each person in the world we create a directed edge only to their ten closest friends (but not to anyone else they know on a first-name basis). In the resulting “closest-friend” version of the social network, is it possible that for each pair of people in the world, there is a path of at most six edges

---

<sup>2</sup>Of course, it can take very large values of  $n$  for this distinction to become truly pronounced; recall Figure 20.6, which showed the results of simulations on networks with 400 million nodes.



# Chapter 23

## Voting

In the previous chapter, we saw a first example of an institution that can synthesize information held by many people, through the ways in which markets serve to aggregate the individual beliefs of investors. We now turn to a second fundamental institution: voting.

### 23.1 Voting for Group Decision-Making

Like markets, voting systems also serve to aggregate information across a group, and as a result, it's hard to draw a perfectly clear dividing line between these two kind of institutions. But there are definite distinctions between the respective settings in which they are typically applied. A first important distinction is that voting is generally used in situations where a group of people is expressly trying to reach a single decision that in a sense will speak for the group. When a population votes on a set of candidates or ballot initiatives, a legislative body votes on whether to pass a bill, a jury votes on a verdict in a trial, a prize committee votes on the recipient of an award, or a group of critics votes on the top movies of the past century, the resulting decision is a single outcome that stands for the group, and has some kind of binding effect going forward. In contrast, markets synthesize the opinions of a group more indirectly, as investors' beliefs are conveyed implicitly through their transactions in the market — choosing how much to invest or to bet, choosing whether to buy or not to buy, and so forth. The overt goal of the market is to enable these transactions, rather than any broader synthesis or group decision that might in fact arise from the transactions in aggregate.

There are other important distinctions as well. A simple but important one is that the choices in a market are often numerical in nature (how much money to transact in various ways), and the synthesis that takes place generally involves arithmetic on these quantities —

---

D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010. Draft version: June 10, 2010.

weighted averages and other measures. Many of the key applications of voting, on the other hand, take place in situations where there's no natural way to "average" the preferences of individuals — since the preferences are over different people, different policy decisions, or different options under a largely subjective criterion. Indeed, as we will see in this chapter, much of the richness of the theory of voting comes from precisely this attempt to combine preferences in the absence of simple metaphors like averaging.

The notion of voting encompasses a broad class of methods for reaching a group decision. For example, the methods to reach a jury verdict, an outcome in a U.S. Presidential Election, or a winner of college football's Heisman Trophy are all distinct, and these distinctions have effects both on the process and the result. Moreover, voting can be used in settings where a single "winner" must be chosen, as well as in situations where the goal is to produce a ranked list. Examples of the latter include the ranking of college sports teams by aggregating multiple polls, or published rankings of the greatest movies, songs, or albums of all time by combining the opinions of many critics.

Voting is often used in situations where the voters disagree because of genuine divergence in their subjective evaluations. For example, film critics who disagree on whether to rank *Citizen Kane* or *The Godfather* as the greatest movie of all time are generally not disagreeing because they lack relevant information about the two movies — we can expect that they are closely familiar with both — but because of differing aesthetic evaluations of them. In other cases, however, voting is used to achieve group decisions where the difficulty is a lack of information — where the members of the group would likely be unanimous if they had all the information relevant to the decision. For example, jury verdicts in criminal trials often hinge on genuine uncertainty as to whether the defendant committed the crime; in such cases one expects that jurors all have approximately the same goal in mind (determining the correct verdict), and the differences are in their access to and processing of the available information. We will consider both of these settings in this chapter.

Ideas from the theory of voting have been adopted in a number of recent on-line applications [140]. Different Web search engines produce different rankings of results; a line of work on *meta-search* has developed tools for combining these rankings into a single aggregate ranking. Recommendation systems for books, music, and other items — such as Amazon's product-recommendation system — have employed related ideas for aggregating preferences. In this case, a recommendation system determines a set of users whose past history indicates tastes similar to yours, and then uses voting methods to combine the preferences of these other users to produce a ranked list of recommendations (or a single best recommendation) for you. Note that in this case, the goal is not a single aggregate ranking for the whole population, but instead an aggregate ranking for each user, based on the preferences of similar users.

Across all of these different contexts in which voting arises, one sees a recurring set of

questions. How should we produce a single ranking from the conflicting opinions provided by multiple voters? Is some version of majority voting a good mechanism? Is there a better one? And ultimately, what does it even mean for a voting system to be good? These are some of the questions we address in this chapter.

## 23.2 Individual Preferences

The goal of a voting system, for our purposes, can be described as follows. A group of people is evaluating a finite set of possible *alternatives*; these alternatives could correspond to political candidates, possible verdicts in a trial, amounts of money to spend on national defense, nominees for an award, or any other set of options in a decision. The people involved wish to produce a single *group ranking* that orders the alternatives from best to worst, and that in some sense reflects the collective opinion of the group. Of course, the challenge will be to define what it means to “reflect” the multiple opinions held by members of the group.

To begin with, let’s consider how to model the opinions of any one member of the group. We suppose that for each individual, he or she is able to determine a preference between any two alternatives when presented with these two as a pair. If individual  $i$  prefers alternative  $X$  to alternative  $Y$ , then we write  $X \succ_i Y$ . (Sometimes, for ease of discussion, we will say that  $X$  “defeats”  $Y$  according to  $i$ ’s preferences.) Thus, for example, if a set of film critics are each given a large list of movies and asked to express preferences, we could write *Citizen Kane*  $\succ_i$  *The Godfather* to express the fact that critic  $i$  prefers the former movie to the latter. We will sometimes refer to an individual’s preferences over all pairs of alternatives, represented by  $\succ_i$ , as this individual’s *preference relation* over the alternatives.

**Completeness and Transitivity.** We require that individual preferences satisfy two properties. The first is that each person’s preferences are *complete*: for each pair of distinct alternatives  $X$  and  $Y$ , either she prefers  $X$  to  $Y$ , or she prefers  $Y$  to  $X$ , but not both. It is possible to extend the theory here to consider the possibility that an individual has ties in her preferences (i.e. for some pairs of alternatives, she likes them equally), and also the possibility that for some pairs of alternatives, an individual has no preference (perhaps because individual  $i$  has no knowledge of one of  $X$  or  $Y$ ). Both of these extensions introduce interesting complications, but for this chapter we focus on the case in which each individual has a preference between each pair of alternatives.

The second requirement is that each individual’s preferences be *transitive*: if an individual  $i$  prefers  $X$  to  $Y$  and  $Y$  to  $Z$ , then  $i$  should also prefer  $X$  to  $Z$ . This seems like a very sensible restriction to impose on preferences, since otherwise we could have situations in which an individual had no apparent favorite alternative. In other words, suppose we were evaluating preferences over flavors of ice cream, and we had an individual  $i$  for whom *Chocolate*  $\succ_i$

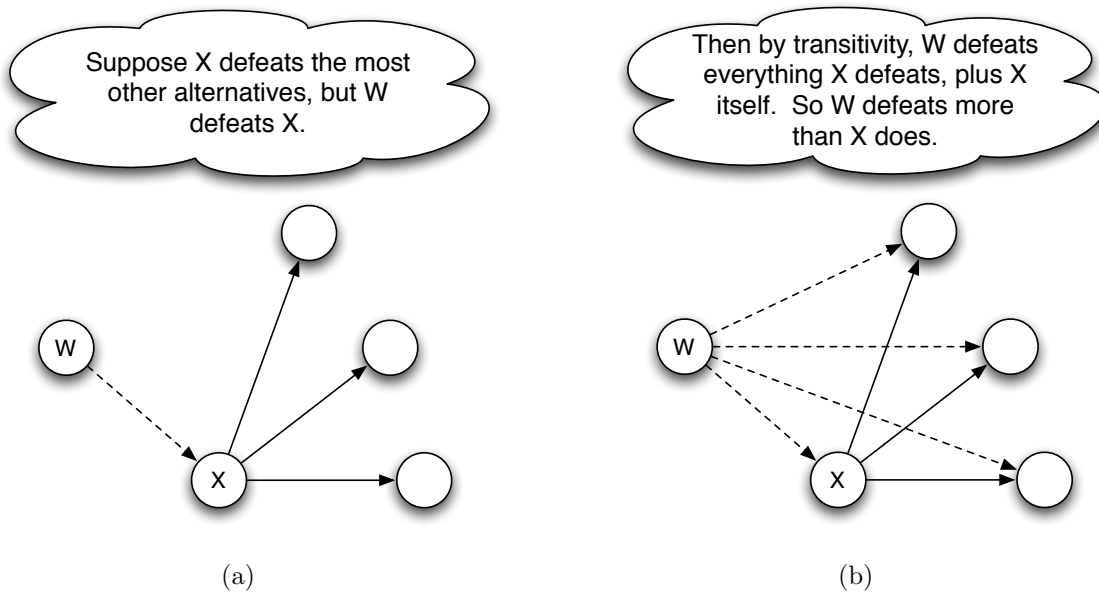


Figure 23.1: With complete and transitive preferences, the alternative  $X$  that defeats the most others in fact defeats all of them. If not, some other alternative  $W$  would defeat  $X$  (as in (a)), but then by transitivity  $W$  would defeat more alternatives than  $X$  does (as in (b)).

*Vanilla*, and *Vanilla*  $\succ_i$  *Strawberry* — and, in a violation of transitivity, also *Strawberry*  $\succ_i$  *Chocolate*. A simple informal argument for why such preferences seem pathological is the following: if individual  $i$  were to walk up to the counter at an ice cream shop and see all three flavors on display, which would she choose? This would in some sense have to be her favorite, despite the fact that each of the three flavors is defeated by some other flavor in her preferences. There has been a long line of work exploring the philosophical and psychological basis for transitive preferences, as well as identifying natural situations in which non-transitive preferences can in fact arise [12, 41, 163]. For our purposes here, we will assume that each individual's preferences are transitive.

**Individual Rankings.** Thus far we have been expressing an individual's opinions about a set of alternatives in terms of his or her preferences over pairs. An alternate model for opinions would be to imagine that each individual produces a completely ranked list of all the alternatives, ranking them from best to worst.

Notice that from such a ranked list, we could define a preference relation  $\succ_i$  very simply: we'd say that  $X \succ_i Y$  if alternative  $X$  comes before alternative  $Y$  in  $i$ 's ranked list. In this case, we'll say that *the preference relation arises from the ranked list*. It's not hard to see that if a preference relation arises from a ranked list of the alternatives, then it must be complete and transitive: completeness holds since for each pair of alternatives, one precedes

the other in the list; and transitivity holds since if  $X$  precedes  $Y$  and  $Y$  precedes  $Z$  in the list, then  $X$  must also precede  $Z$ .

What is somewhat less obvious is that this fact holds in the opposite direction as well:

*If a preference relation is complete and transitive, then it arises from some ranked list of the alternatives.*

The way to see why this is true is to consider the following method for constructing a ranked list from a complete and transitive preference relation. First, we identify the alternative  $X$  that defeats the most other alternatives in pairwise comparisons — that is, the  $X$  so that  $X \succ_i Y$  for the most other choices of  $Y$ . We claim that this  $X$  in fact defeats all the other alternatives:  $X \succ_i Y$  for all other  $Y$ .

We'll see why this is true in a moment; but first, let's see why this fact lets us construct the ranked list we want. To begin with, having established that  $X$  defeats all other alternatives, we can safely put it at the front of the ranked list. We now remove  $X$  from the set of alternatives, and repeat exactly the same process on the remaining alternatives. The preferences defined by  $\succ_i$  are still complete and transitive on the remaining alternatives, so we can apply our claim again on this smaller set: for the alternative  $Y$  that defeats the most others in this set, it in fact defeats every remaining alternative. Hence  $Y$  defeats every alternative in the original set except for  $X$ , so we can put  $Y$  second in the list, remove it too from the set of alternatives, and continue in this way until we exhaust the finite set of alternatives. The way we've constructed the list, each alternative is preferred to all the alternatives that come after it, and so  $\succ_i$  arises from this ranked list.

All of this depends on showing that for any complete and transitive preferences over a set of alternatives (including the original preferences, and the ones we get as we remove alternatives), the alternative  $X$  that defeats the most others in fact defeats all of them. Here is an argument showing why this is true (also illustrated in Figure 23.1). We suppose, for the sake of a contradiction, that it were not true; then there would be some alternative  $W$  that defeats  $X$ . But then, for every  $Y$  that is defeated by  $X$ , we'd have  $W \succ_i X$  and  $X \succ_i Y$ , and so by transitivity  $W \succ_i Y$ . The conclusion is that  $W$  would defeat everything  $X$  defeats, and also defeat  $X$  — so  $W$  would defeat more alternatives than  $X$  does. This is a contradiction, since we chose  $X$  as the alternative that defeats the most others. Therefore, our assumption that some  $W$  defeats  $X$  cannot be correct, and so  $X$  in fact defeats all other alternatives. This argument justifies our construction of the ranked list.

In view of all this, when we have a complete and transitive preference relation, we can equally well view it as a ranked list. Both of these views will be useful in the discussion to follow.

## 23.3 Voting Systems: Majority Rule

In the previous section we developed a way to talk about the individual preference relations that we’re seeking to combine. We can now describe a *voting system* (also called an *aggregation procedure*) as follows: it is any method that takes a collection of complete and transitive individual preference relations — or equivalently, a collection of individual rankings — and produces a *group ranking*.

This is a very general definition, and at this level of generality it may be hard to see what makes for a “reasonable” voting system. So we begin, in this section and the next, by discussing two of the most common classes of voting systems. By considering these, we’ll start to identify some of the principles — and pathologies — at work in voting more generally.

**Majority Rule and the Condorcet Paradox.** When there are only two alternatives, the most widely used voting system — and arguably the most natural — is *majority rule*. Under majority rule, we take the alternative that is preferred by a majority of the voters and rank it first, placing the other alternative second. For this discussion we will assume that the number of voters is odd, so that we won’t have to worry about the possibility of majority rule producing ties.

Since majority rule is so natural in the case of two alternatives, it is natural to try designing a voting system based on majority rule when there are more than two alternatives. This, however, turns out to be remarkably tricky. Probably the most direct approach is to first create *group preferences*, by applying majority rule to each pair of alternatives, and then trying to turn these group preferences into a group ranking. That is, we create a group preference relation  $\succ$  out of all the individual preferences  $\succ_i$  as follows. For each pair of alternatives  $X$  and  $Y$ , we count the number of individuals for whom  $X \succ_i Y$  and the number of individuals for whom  $Y \succ_i X$ . If the first number is larger than the second, then we say that the group preference  $\succ$  satisfies  $X \succ Y$ , since a majority of the voters prefer  $X$  to  $Y$  when these two alternatives are considered in isolation. Similarly, we say  $Y \succ X$  in the group preference if  $Y \succ_i X$  for a majority of the individuals  $i$ . Since the number of voters is odd, we can’t have equal numbers favoring  $X$  and favoring  $Y$ . So for every distinct pair of alternatives we will have exactly one of  $X \succ Y$  or  $Y \succ X$ . That is, the group preference relation is complete.

There’s no problem getting this far; the surprising difficulty is that *the group preferences may not be transitive, even when each individual’s preferences are transitive*. To see how this can happen, suppose that we have three individuals named 1, 2, and 3, and three alternatives named  $X$ ,  $Y$ , and  $Z$ . Suppose further that individual 1’s ranking is

$$X \succ_1 Y \succ_1 Z, \tag{23.1}$$

College	National Ranking	Average Class Size	Scholarship Money Offered
X	4	40	\$3000
Y	8	18	\$1000
Z	12	24	\$8000

Figure 23.2: When a single individual is making decisions based on multiple criteria, the Condorcet Paradox can lead to non-transitive preferences. Here, if a college applicants wants a school with a high ranking, small average class size, and a large scholarship offer, it is possible for each option to be defeated by one of the others on a majority of the criteria.

individual 2's ranking is

$$Y \succ_2 Z \succ_2 X, \quad (23.2)$$

and individual 3's ranking is

$$Z \succ_3 X \succ_3 Y. \quad (23.3)$$

Then using majority-rule to define group preferences, we'd have  $X \succ Y$  (since  $X$  is preferred to  $Y$  by both 1 and 3),  $Y \succ Z$  (since  $Y$  is preferred to  $Z$  by both 1 and 2), and  $Z \succ X$  (since  $Z$  is preferred to  $X$  by both 2 and 3). This violates transitivity, which would require  $X \succ Z$  once we have  $X \succ Y$  and  $Y \succ Z$ .

The possibility of non-transitive group preferences arising from transitive individual preferences is called the *Condorcet Paradox*, after the Marquis de Condorcet, a French political philosopher who discussed it in the 1700s. And there's something genuinely counter-intuitive about it. If we recall our earlier discussion of non-transitive preferences as being somehow "incoherent," the Condorcet Paradox describes a simple scenario in which a set of people, each with perfectly plausible preferences, manages to behave incoherently when forced to express their collective preferences through majority rule. For example, let's return to our example of an individual who prefers Chocolate to Vanilla to Strawberry to Chocolate. Even if we were to assume that no one individually behaves this way, the Condorcet Paradox shows how this can arise very naturally as the group preferences of a set of ice-cream eating friends, when they plan to share a pint of ice cream and decide on which flavor to buy using majority rule.

The Condorcet Paradox has in fact also been used to show how a single person can naturally be led to form non-transitive individual preferences [41, 163]. Consider, for example, a student deciding which college to attend. She prefers to go to a college that is highly ranked, that has a small average class size, and that offers her a significant amount in scholarship money. Suppose she has been admitted to the following three colleges, with characteristics as described in Figure 23.2.

In comparing colleges, the student was planning to decide between pairs of colleges by favoring the one that did better on a majority of these three criteria. Unfortunately, this

leads to the preferences  $X \succ_i Y$  (since  $X$  is better than  $Y$  on ranking and scholarship money),  $Y \succ_i Z$  (since  $Y$  is better than  $Z$  on ranking and average class size), and  $Z \succ_i X$  (since  $Z$  is better than  $X$  on average class size and scholarship money). It's not hard to see the analogy: each criterion is like a voter, and the student's "individual preference relation" is really the group preference relation synthesized from these three criteria. But it does show some of the complications that arise even when one individual engages in decision-making in the presence of multiple criteria.

**Voting Systems based on Majority Rule.** The Condorcet Paradox portends trouble for the design of voting systems in general, but given that we need some way to produce an actual group ranking (including an actual top-ranked alternative), it's still worth exploring what can be done using majority rule. We'll focus on methods for selecting a top-ranked alternative, which we'll think of as the "group favorite"; to produce a full ranked list, one could first select a group favorite, remove it from the available alternatives, and then apply the procedure repeatedly to what's left.

One natural approach for finding a group favorite is as follows. We arrange all the alternatives in some order, and then eliminate them one-by-one in this order using majority rule. Thus, we compare the first two alternatives by majority vote, compare the winner of this vote to the third alternative, then compare the winner of that to the fourth alternative, and so on. The winner of the final comparison is deemed to be the group favorite. We can represent this in a pictorial way as in Figure 23.3(a), showing the sequence of eliminations in a four-alternative example, with alternatives  $A$  and  $B$  compared first, then the winner compared to  $C$ , and the winner of that compared to  $D$ . We can think of this as an agenda for a meeting in which pairs of alternatives are proposed to the group, majority votes are taken, and a group favorite emerges from this.

This is an example of a more general strategy for using majority rule over pairs of alternatives to find a group favorite: we can arrange them in any kind of "elimination tournament," in which alternatives are paired off against each other in some fashion, with the winner advancing to a subsequent round while the loser is eliminated. The alternative that eventually emerges as the overall winner of this tournament is declared to be the group favorite. The system we were just discussing, shown in Figure 23.3(a), is one such way to structure an elimination tournament; Figure 23.3(b) depicts another.

**Pathologies in Voting Systems based on Majority Rule.** These systems do produce a group favorite (and, by repeatedly invoking the system on the remaining alternatives, also produce a group ranking). The Condorcet Paradox, however, can be used to uncover an important pathology that such systems exhibit: their outcomes are susceptible to a kind of *strategic agenda-setting*. Let's go back to our original example of the Condorcet Paradox,



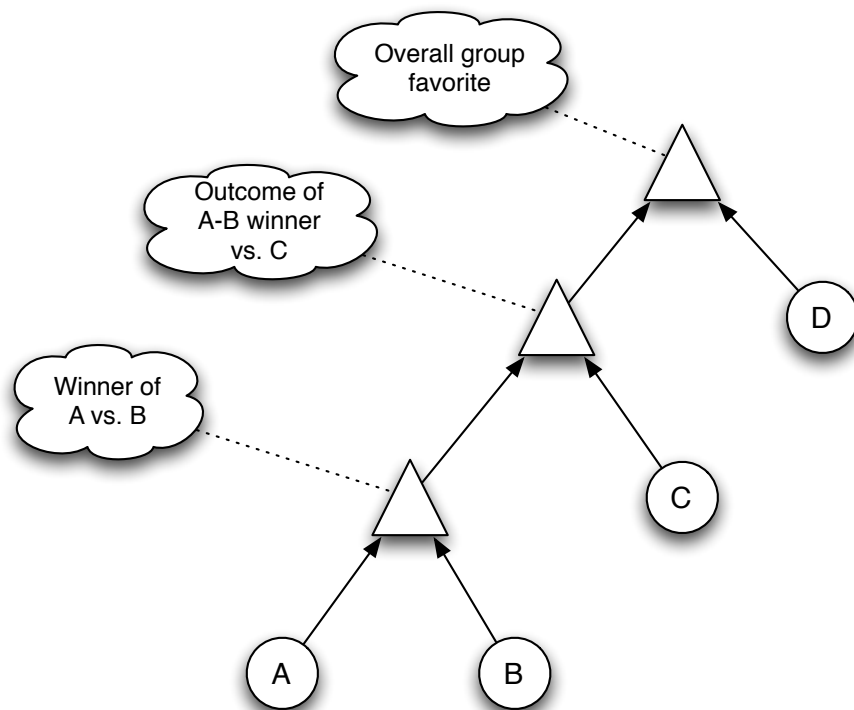
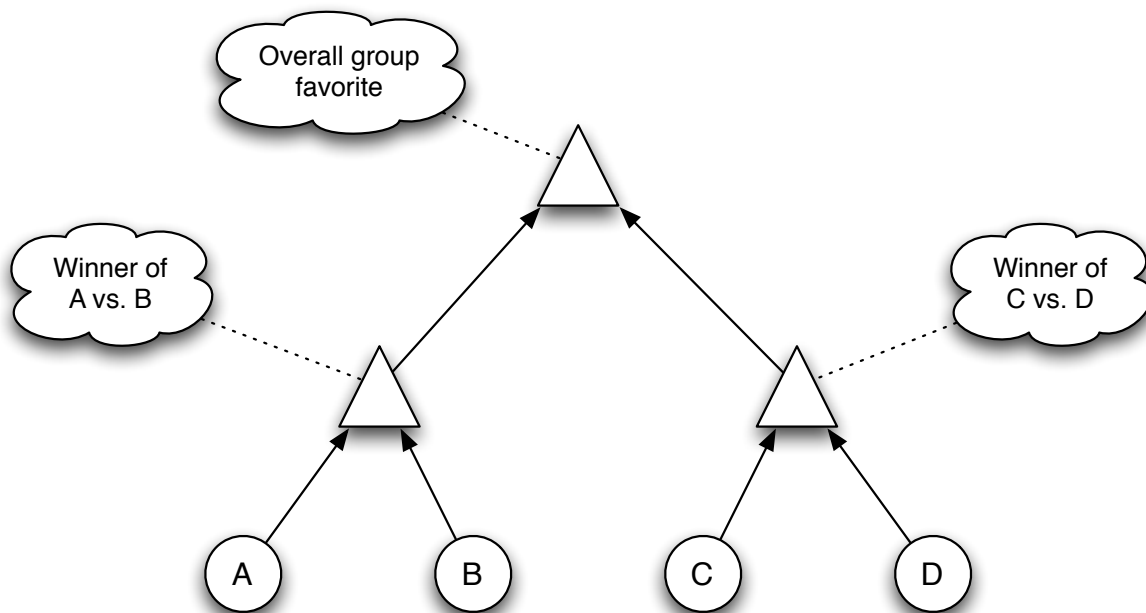
(a) *Introducing new alternatives one at a time.*(b) *Pairing off alternatives in a different order.*

Figure 23.3: One can use majority rule for pairs to build voting systems on three or more alternatives. The alternatives are considered according to a particular “agenda” (in the form of an elimination tournament), and they are eliminated by pairwise majority vote according to this agenda. This produces an eventual winner that serves as the overall group favorite.

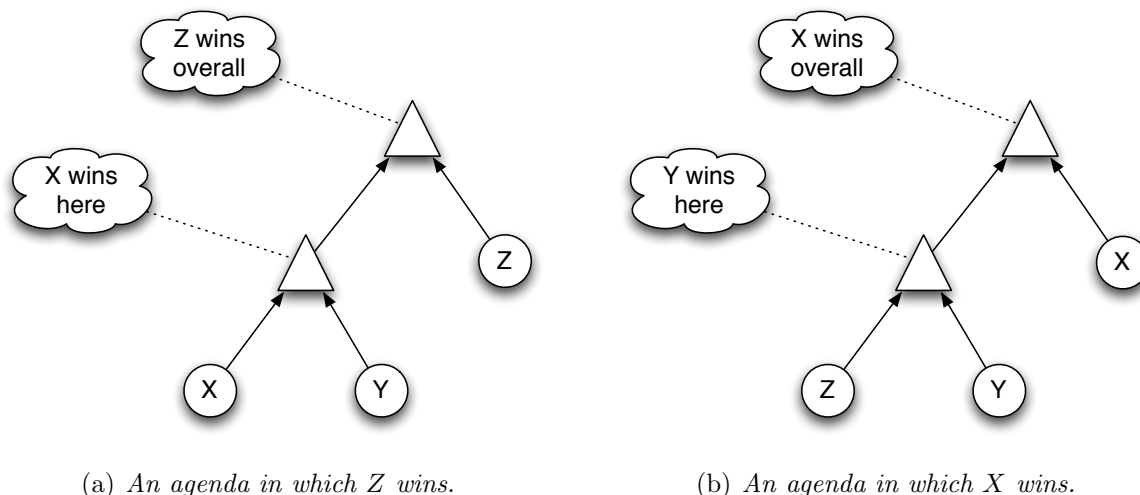


Figure 23.4: With individual rankings as in the Condorcet Paradox, the winner of the elimination tournament depends entirely on how the agenda is set.

where three voters had the individual rankings over alternatives  $X$ ,  $Y$ , and  $Z$  given by lists (23.1)–(23.3). They decide to choose a group favorite using a version of the system shown in Figure 23.3(a), scaled down to three alternatives: they will first perform majority vote between two of the alternatives, and then perform majority vote between the winner of this first vote and the remaining alternative.

The question then becomes how to set the agenda for this process. That is, which two of the alternatives  $X$ ,  $Y$ , and  $Z$  will be voted on first, and which one will be held until the final vote? Because of the structure of the individual preferences, the choice of agenda in this case has a decisive effect on the outcome. If, as in Figure 23.4(a), alternatives  $X$  and  $Y$  are paired off first, then  $X$  will win this first vote but then be defeated by  $Z$  as the group favorite. On the other hand, if alternatives  $Y$  and  $Z$  are paired off first (as in Figure 23.4(b)), then  $Y$  will win this first vote but then be defeated by  $X$  as the group favorite. (We could do a similar thing to have  $Y$  be the group favorite.)

So with individual preferences as in the Condorcet paradox, the overall winner is determined entirely by how the votes between pairs are sequenced. To put it a different way, if the voter who likes  $Z$  best gets to set the agenda, then she can sequence the votes so  $Z$  wins; but if the voter who likes  $X$  or  $Y$  best gets to set the agenda, then he can sequence the votes so his respective favorite wins. The group favorite is thus determined by the individual who controls the agenda. Nor can this be remedied by using a system in which voters can re-introduce alternatives once they've been eliminated: with preferences as in the Condorcet Paradox, there's always an alternative that can be re-introduced to defeat the current candidate for the group favorite, and so a process in which alternatives can be re-introduced for

consideration would never come to an end.

Earlier in this section — using an example of a student choosing colleges based on multiple criteria, in Figure 23.2 — we observed that the Condorcet Paradox can also capture pathologies in the decisions made by a single individual, rather than a group. The problem of agenda-setting has an analogue in the context of such individual decisions as well. Suppose, for example, that the student in our earlier college-choice example makes the natural decision to eliminate choices one at a time as acceptance offers come in. Then if the acceptances arrive in the order  $X, Y, Z$ , she will eliminate  $Y$  in favor of  $X$  when  $Y$  arrives ( $X$  has a higher ranking and a higher scholarship offer) and then eliminate  $X$  in favor of  $Z$  when the acceptance from  $Z$  arrives (since  $Z$  has a smaller average class size and a higher scholarship offer). Each of these makes sense as a pairwise decision, and it leads to  $Z$  as her overall decision — but it has the property that  $Y$ , which she eliminated first, is in fact a choice she preferred to  $Z$ . This is precisely the problem of having a final decision that depends on the agenda by which alternatives are considered.

## 23.4 Voting Systems: Positional Voting

A different class of voting systems tries to produce a group ranking directly from the individual rankings, rather than building up the group ranking from pairwise comparisons of alternatives. In this type of system, each alternative receives a certain *weight* based on its positions in all the individual rankings, and the alternatives are then ordered according to their total weight. A simple example of such a system is the *Borda Count*, named for Jean-Charles de Borda, who proposed it in 1770. The Borda Count is often used to choose the winners of sports awards, such as the Heisman trophy in college football; a variant of it is used to select the Most Valuable Player in professional baseball; and it is used by the Associated Press and United Press International to rank sports teams.

In the Borda Count, if there are  $k$  alternatives in total, then individual  $i$ 's ranking confers a weight of  $k - 1$  on her first-ranked alternative, a weight of  $k - 2$  on her second-ranked alternative, and so on down to a weight of 1 on her second-to-last alternative, and a weight of 0 on her last alternative. In other words, each alternative receives a weight from individual  $i$  equal to the number of other alternatives ranked lower by  $i$ . The total weight of each alternative is simply the sum of the weights it receives from each of the individuals. The alternatives are then ordered according to their total weights. (We will suppose that if two alternatives receive the same total weight, then some tie-breaking system arranged in advance is used to decide which of these two alternatives to place in front of the other.)

For example, suppose there are four alternatives, named  $A, B, C$ , and  $D$ , and there are two voters with the individual rankings

$$A \succ_1 B \succ_1 C \succ_1 D$$

and

$$B \succ_2 C \succ_2 A \succ_2 D.$$

Then the weight assigned by the Borda Count to alternative  $A$  is  $3 + 1 = 4$ , the weight assigned to  $B$  is 5, the weight assigned to  $C$  is 3, and the weight assigned to  $D$  is 0. Therefore, sorting the weights in descending order, the group ranking is

$$B \succ A \succ C \succ D.$$

It is easy to create variants of the Borda Count that retain its basic flavor: we can assign any number of “points” to each position in each list, and then rank the alternatives by the total number of points they receive based on their positions in all lists. The Borda Count assigns  $k - 1$  points for first,  $k - 2$  points for second, and so forth, but one could imagine versions that assign points differently: for example, to make only the top three positions in each individual ranking matter, one could assign 3 points for first, 2 for second, 1 for third, and 0 points for all other positions, with the group ranking still determined by the total number of points the alternatives receive. We refer to any system of this type as a *positional voting system*, since the alternatives receive numerical weights based on their positions on the individual rankings.

A key appealing feature of the Borda Count is that — ignoring ties — it always produces a complete, transitive ranking for a set of alternatives. This is simply by its definition, since it creates a single numerical criterion along which to sort the alternatives (including, as noted previously, a rule for tie-breaking). But the Borda Count also has some fundamental pathologies, as we now discuss.

**Pathologies in Positional Voting Systems.** Most of the problems with the Borda Count, and with positional voting systems more generally, arise from the fact that competition for top spots in the group ranking can depend critically on the rankings of alternatives that are further down in the list.

Here’s a hypothetical scenario illustrating how this can happen. Suppose that a magazine writes a column in which it asks five film critics to discuss their choice for the greatest movie of all time; the two movies discussed in the column are *Citizen Kane* and *The Godfather*, and the column ends with a majority-vote decision on the winner. Critics 1, 2, and 3 favor *Citizen Kane*, while critics 4 and 5 favor *The Godfather*.

At the last minute, however, the editors decide the column needs a more “modern” feel, so they introduce *Pulp Fiction* as a third option that needs to be discussed and evaluated. Since there are now three options, the magazine decides to have each critic produce a ranking, and then use the Borda Count for the overall decision that will serve as the punchline of the column. The first three critics (who all prefer older movies) each report the ranking

$$\textit{Citizen Kane} \succ_i \textit{The Godfather} \succ_i \textit{Pulp Fiction}.$$

Critics 4 and 5 (who only like movies made in the last 40 years) each report the ranking

$$\textit{The Godfather} \succ_i \textit{Pulp Fiction} \succ_i \textit{Citizen Kane}.$$

Applying the Borda Count, we see that *Citizen Kane* receives a weight of 2 from each of the first three critics, and a weight of 0 from the last two, for a total of 6. *The Godfather* receives a weight of 1 from each of the first three critics, and a weight of 2 from the last two, for a total of 7. *Pulp Fiction* receives a weight of 0 from each of the first three critics, and a weight of 1 from the last two, for a total of 2. As a result, the Borda Count produces *The Godfather* as the overall group favorite.

Notice what's happened here. The outcome of the head-to-head comparison between *Citizen Kane* and *The Godfather* remains the same as before — *Citizen Kane* is favored by a vote of three to two. But because a third alternative was introduced, the identity of the group favorite has changed. Moreover, this is not because the group was particularly fond of this new, third alternative — the third alternative loses in a head-to-head vote against *each* of the two existing alternatives. To put the difficulty in another way: *Citizen Kane* fails to rank first in the Borda Count even though it defeats each of the other two alternatives in a head-to-head comparison under majority rule. So what we find is that the outcome in the Borda Count can depend on the presence of alternatives that intuitively seem “irrelevant” — weak alternatives that essentially act as “spoilers” in shifting the outcome from one higher-ranked alternative to another.

The possibility of such a result suggests further difficulties with the Borda Count — specifically, the problem of *strategic misreporting of preferences*. To see how this happens, let's consider a slightly different scenario. Suppose in our previous story that critics 4 and 5 actually had the true ranking

$$\textit{The Godfather} \succ_i \textit{Citizen Kane} \succ_i \textit{Pulp Fiction}.$$

In other words, in this version of the story, all five critics agree that *Pulp Fiction* should be ranked last among these three movies. If we were to run the Borda Count on this set of five individual rankings, the group ranking would place *Citizen Kane* first (it would receive a total weight of  $3 \cdot 2 + 2 \cdot 1 = 8$  to *The Godfather's*  $3 \cdot 1 + 2 \cdot 2 = 7$ ). However, suppose that critics 4 and 5 understand the pathologies that are possible with the Borda Count, and they decide in advance to misreport their rankings as

$$\textit{The Godfather} \succ_i \textit{Pulp Fiction} \succ_i \textit{Citizen Kane}.$$

Then we have the individual rankings from the previous scenario, and *The Godfather* ends up being ranked first.

The underlying point is that voters in the Borda Count can sometimes benefit by lying about their true preferences, particularly so as to downgrade the overall group ranking of an alternative that many other voters will put at the top of their individual rankings.

**Examples from U.S. Presidential Elections.** Versions of these pathologies are familiar from U.S. Presidential elections as well. The full Presidential Election process in the United States has a complex specification, but if we think about how states choose their electors in the general election — i.e. how they choose which candidate will receive the state’s electoral votes — it is generally done using *plurality voting*: the candidate who is top-ranked by the most voters wins. (The U.S. Constitution doesn’t require this, and some states have considered other methods and used others in the past, but this is the typical system.)

If we think about it, plurality voting is in fact a positional voting system, since an equivalent way to run it is as follows. We ask each voter to report an individual ranking of all the candidates. Each individual ranking then confers a weight of 1 to the candidate at the top of the ranking, and a weight of 0 to all the other candidates. The candidate with the greatest total weight from these rankings is declared the winner. Note that this is just a different way of saying that the candidate who is top-ranked by the most voters wins, but it makes it clear that this system fits the structure of a positional method.

Plurality voting exhibits difficulties analogous to what we observed for the Borda Count. With only two candidates, plurality voting is the same as majority rule; but with more than two candidates, one sees recurring “third-party” effects, where an alternative that is the favorite of very few people can potentially shift the outcome from one of the two leading contenders to the other. In turn, this causes some voters to make their choices strategically, misreporting their top-ranked choice so as to favor a candidate with a better chance of winning. Such issues have been present in recent U.S. Presidential elections, and their effects in important earlier elections, such as the election of Abraham Lincoln in 1860, have also been studied [384].

## 23.5 Arrow’s Impossibility Theorem

We have now looked at a number of different voting systems, and we’ve seen that when there are more than two alternatives under consideration, they all exhibit pathological behavior. If we were to consider further voting systems used in practice, we’d find they too suffered from inherent problems in the way they produce a group ranking. At some point, however, it makes sense to step back from specific voting systems and ask a more general question: is there *any* voting system that produces a group ranking for three or more alternatives, and avoids all of the pathologies we’ve seen thus far?

Making this question concrete requires that we precisely specify all the relevant definitions. We’ve already discussed the precise definition of a voting system: for a fixed number of voters  $k$ , it is any function that takes a set of  $k$  individual rankings and produces a group ranking. The other thing we need to do is to specify what it means for a voting system to be free of pathologies. We will do this by specifying two properties that we would like a

reasonable voting system to satisfy. They are the following.

- First, if there is any pair of alternatives  $X$  and  $Y$  for which  $X \succ_i Y$  in the rankings of every individual  $i$ , then the group ranking should also have  $X \succ Y$ . This is a very natural condition, known as the *Pareto Principle*, or *Unanimity*; it simply requires that if everyone prefers  $X$  to  $Y$ , then the group ranking should reflect this. One can think of Unanimity as ensuring that the group ranking be responsive to the individual rankings in at least a minimal way.
- Second, we require that for each pair of alternatives, the ordering of  $X$  and  $Y$  in the group ranking should depend only on how each individual ranks  $X$  and  $Y$  relative to each other. In other words, suppose we have a set of individual rankings that produces a group ranking in which  $X \succ Y$ . If we then take some third alternative  $Z$  and shift its position in some of the individual rankings, while leaving the relative ordering of  $X$  and  $Y$  unchanged, then the voting system should still produce  $X \succ Y$  for this new set of individual rankings.

This condition is called *Independence of Irrelevant Alternatives (IIA)*, since it requires that the group ranking of  $X$  and  $Y$  should depend only on voter preferences between  $X$  and  $Y$ , not on how they evaluate other alternatives. IIA is more subtle than Unanimity, but the failure of IIA is in fact responsible for most of the pathological behavior we saw in our earlier discussions of specific voting systems. It was clearly at work in the strategic misreporting of preferences for the Borda Count, since there the shift in ranking of a third alternative  $Z$  was sufficient to change the outcome between two other alternatives  $X$  and  $Y$ . It also plays a role in the problem of strategic agenda-setting for elimination systems based on majority rule: the key idea there was to choose an agenda that eliminated one alternative  $X$  early, before it could be paired against an alternative  $Y$  that it would in fact defeat.

**Voting Systems that Satisfy Unanimity and IIA.** Since Unanimity and IIA are both reasonable properties, it's natural to ask what voting systems satisfy them. When there are only two alternatives, majority rule clearly satisfies both: it favors  $X$  to  $Y$  when all voters do, and — since there are only two alternatives — the group ranking of  $X$  and  $Y$  clearly does not depend on any other alternatives.

When there are three or more alternatives, it's trickier to find a voting system that satisfies these two properties: neither the positional systems nor the systems based on majority rule that we've considered will work. There is, however, a voting system that satisfies the two properties: dictatorship. That is, we pick one of the individuals  $i$ , and we simply declare the group ranking to be equal to the ranking provided by individual  $i$ . Notice that there are really  $k$  different possible voting systems based on dictatorship — one in which each of the

$k$  possible voters is chosen as the dictator.

We can easily check that each of these  $k$  dictatorship systems satisfies Unanimity and IIA. First, if everyone prefers  $X$  to  $Y$ , then the dictator does, and hence the group ranking does. Second, the group ranking of  $X$  and  $Y$  depends only how the dictator ranks  $X$  and  $Y$ , and does not depend on how any other alternative  $Z$  is ranked.

**Arrow’s Theorem.** In the 1950s, Kenneth Arrow proved the following remarkable result [22, 23], which clarifies why it’s so hard to find voting systems that are free of pathological behavior.

*Arrow’s Theorem: If there are at least three alternatives, then any voting system that satisfies both Unanimity and IIA must correspond to dictatorship by one individual.*

In other words, dictatorship is the only voting system that satisfies both Unanimity and IIA.

Since dictatorship is generally viewed as an undesirable property too, Arrow’s Theorem is often phrased as an impossibility result. That is, suppose we say a voting system satisfies *non-dictatorship* if there is no individual  $i$  for which the group ranking always coincides with  $i$ ’s ranking. Then we can phrase Arrow’s Theorem as follows.

*Arrow’s Theorem (equivalent version): If there are at least three alternatives, then there is no voting system that satisfies Unanimity, IIA, and non-dictatorship.*

Ultimately, what Arrow’s Theorem shows us is not that voting is necessarily “impossible,” but that it is subject to unavoidable trade-offs — that any system we choose will exhibit certain forms of undesirable behavior. It therefore helps to focus discussions of voting on how to manage these trade-offs, and to evaluate different voting systems in light of them.

## 23.6 Single-Peaked Preferences and the Median Voter Theorem

Condorcet’s Paradox and Arrow’s Theorem are facts of nature; we cannot make them go away. However, a common approach when faced with an impossibility result is to consider reasonable special cases of the problem where the underlying difficulties do not arise. There has been a long line of research in voting that follows this direction.

The starting point for this line of research is the observation that there’s something a bit unusual about the individual rankings used in the set-up of the Condorcet Paradox. Recall that with three alternatives  $X$ ,  $Y$ , and  $Z$ , and three voters 1, 2, and 3, we had

$$X \succ_1 Y \succ_1 Z$$



$$Y \succ_2 Z \succ_2 X$$

$$Z \succ_3 X \succ_3 Y$$

Suppose that  $X$ ,  $Y$ , and  $Z$  correspond to amounts of money to spend on education or national defense, with  $X$  corresponding to a small amount,  $Y$  to a medium amount, and  $Z$  to a large amount. Then the preferences of voter 1 make sense: she is happiest with the smallest amount, and second-happiest with a medium amount. The preferences of voter 2 also make sense: he is happiest with a medium amount, but if not medium, then he prefers a large amount. The preferences of voter 3, on the other hand, are harder to justify in a simple way: he prefers a large amount, but his second choice is a small amount, with medium coming last. In other words, the first two voters have preferences that can be explained by proximity to a fixed number: each of them has an “ideal” amount that they’d like, and they evaluate the alternatives by how close they come to this ideal. The third voter’s preferences can’t be explained this way: there’s no “ideal” quantity such that both large and small are close to it, but medium isn’t. This is not to say that a person couldn’t hold these preferences (e.g. “if we’re not willing to invest enough in education to do it right, we shouldn’t spend anything at all”), but they’re more unusual.

Similar reasoning would apply if  $X$ ,  $Y$ , and  $Z$  were political candidates arranged in order on a political spectrum, with  $X$  the liberal candidate,  $Y$  the moderate candidate, and  $Z$  the conservative candidate. In this case, voter 1 prefers more liberal candidates; voter 2 prefers moderates and leans conservative when forced to choose between extremes; but voter 3 favors the conservative candidate, followed next by the liberal candidate, with the moderate candidate last. Again, the preferences of voters 1 and 2 can be explained by assuming that each evaluates candidates by their proximity to a personal “ideal” point on the political spectrum, but voter 3’s preferences are less natural in that they can’t be explained this way.

We now describe a way to formalize the “unusualness” in voter 3’s ranking, and we then show that for rankings that do not contain this structure, the Condorcet Paradox cannot arise.

**Single-peaked preferences.** For alternatives corresponding to numerical quantities or linear orderings like a political spectrum, it is reasonable to assume that individual preferences tend to look like the preferences of voters 1 and 2 in our example: each has a particular favorite point in the range of alternatives, and they evaluate alternatives by their proximity to this point. In fact, for our discussion here, it is enough to assume something weaker: simply that each voter’s preferences “fall away” consistently on both sides of their favorite alternative.

To make this precise, let’s assume that the  $k$  alternatives are named  $X_1, X_2, \dots, X_k$ , and that voters all perceive them as being arranged in this order. (Again, we’ll think of the examples of numerical quantities or candidates on a political spectrum.) We say that a

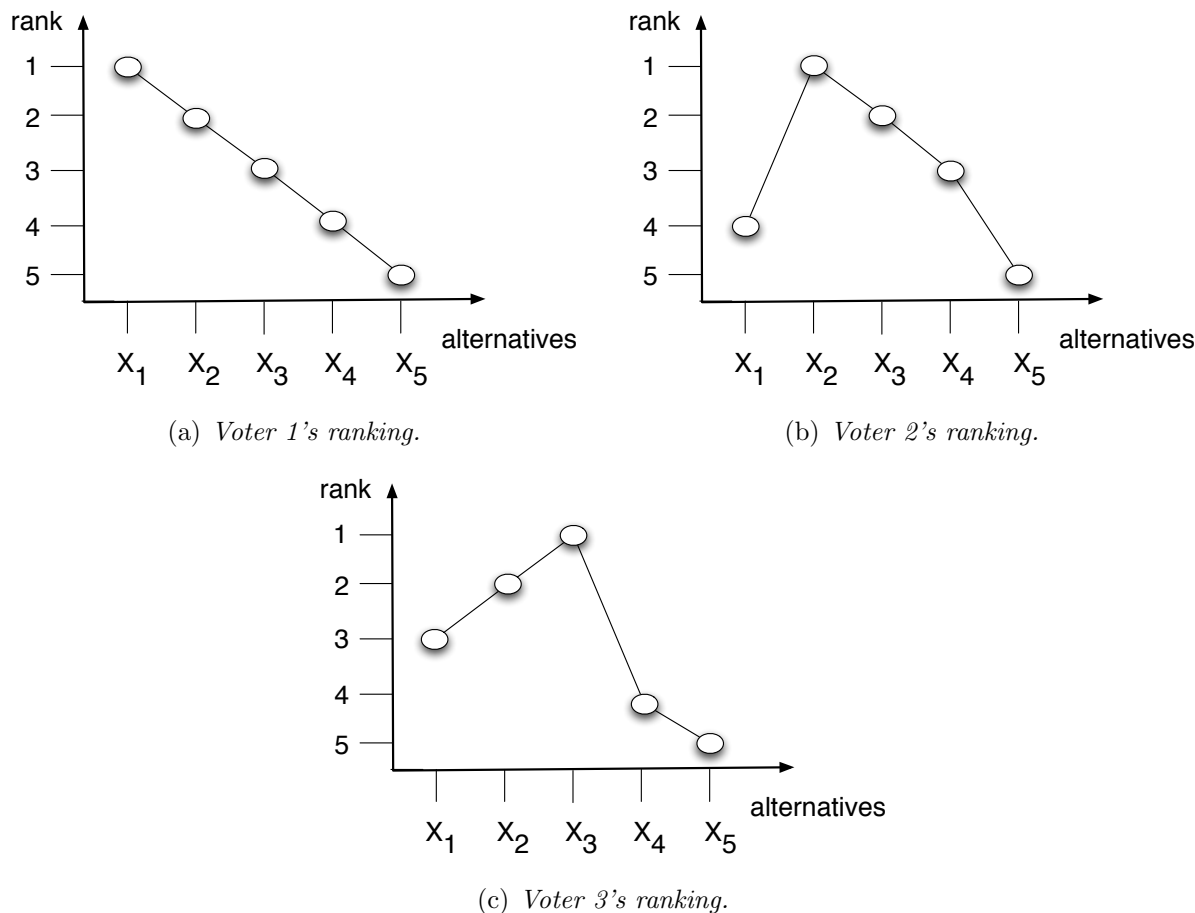


Figure 23.5: With single-peaked preferences, each voter's ranking of alternatives decreases on both sides of a "peak" corresponding to her favorite choice.

voter has *single-peaked preferences* if there is no alternative  $X_s$  for which both neighboring alternatives  $X_{s-1}$  and  $X_{s+1}$  are ranked above  $X_s$ . In other words, a voter never prefers two options that lie on opposite sides of a middle option. (Since we are assuming voters have complete and transitive preferences, we will also refer to single-peaked preferences as single-peaked rankings.)

Such preferences are called single-peaked because the condition we impose is equivalent to the following one: each voter  $i$  has a top-ranked option  $X_t$ , and her preferences fall off on both sides of  $X_t$ :

$$X_t \succ_i X_{t+1} \succ_i X_{t+2} \succ_i \cdots$$

and

$$X_t \succ_i X_{t-1} \succ_i X_{t-2} \succ_i \cdots$$

Pictorially, this can be represented as in Figure 23.5. The example shown there has three

voters with preferences

$$X_1 \succ_1 X_2 \succ_1 X_3 \succ_1 X_4 \succ_1 X_5$$

$$X_2 \succ_2 X_3 \succ_2 X_4 \succ_2 X_1 \succ_2 X_5$$

$$X_3 \succ_3 X_2 \succ_3 X_1 \succ_3 X_4 \succ_3 X_5$$

and each of the three plots shows one of these sets of individual preferences: In the plots, there is an oval for each alternative, and its height corresponds to its position in the list. As drawn, the single peak in an individual's ranking emerges visually as a peak in the plot.

**Majority Rule with Single-Peaked Preferences.** Single-peaked preferences are natural as a model for many kinds of rankings, but their significance in the theory of voting lies in the following observation, made by Duncan Black in 1948 [61].

Recall our first, most basic attempt at synthesizing a group ranking from a set of individual rankings, back in Section 23.3: we would compare each pair of alternatives  $X$  and  $Y$  to each other, using majority rule to produce a group preference of the form  $X \succ Y$  or  $Y \succ X$  (depending on which alternative is preferred by more voters). As before, we'll suppose that the number of voters is odd, so that we don't have to worry about the possibility of ties. Our hope was that the resulting group preference relation  $\succ$  would be complete and transitive, so that we could produce a group ranking from it. Unfortunately, the Condorcet Paradox showed that this hope was in vain: transitive individual preferences can give rise to group preferences that are non-transitive.

But here's the point of the framework we've developed in this section: with single-peaked preferences, our original plan works perfectly. This is the content of the following result.

*Claim: If all individual rankings are single-peaked, then majority rule applied to all pairs of alternatives produces a group preference relation  $\succ$  that is complete and transitive.*

It is not initially clear why this striking fact should be true, but in fact it follows for an intuitively natural reason, as we now describe.

**The Median Individual Favorite.** As in other attempts to construct group rankings, we start by figuring out how to identify a group favorite — an alternative that can be placed at the top of the ranking — and then proceed to fill in further slots of the ranking. Finding a group favorite is the crux of the problem, since that requires us to identify an alternative that defeats every other alternative in a pairwise majority vote.

Let's consider the top-ranked alternative for each voter, and sort this set of individual favorites from left to right, along our linear order. Notice that if several voters have the same alternative as their respective individual favorite, then this alternative will appear multiple

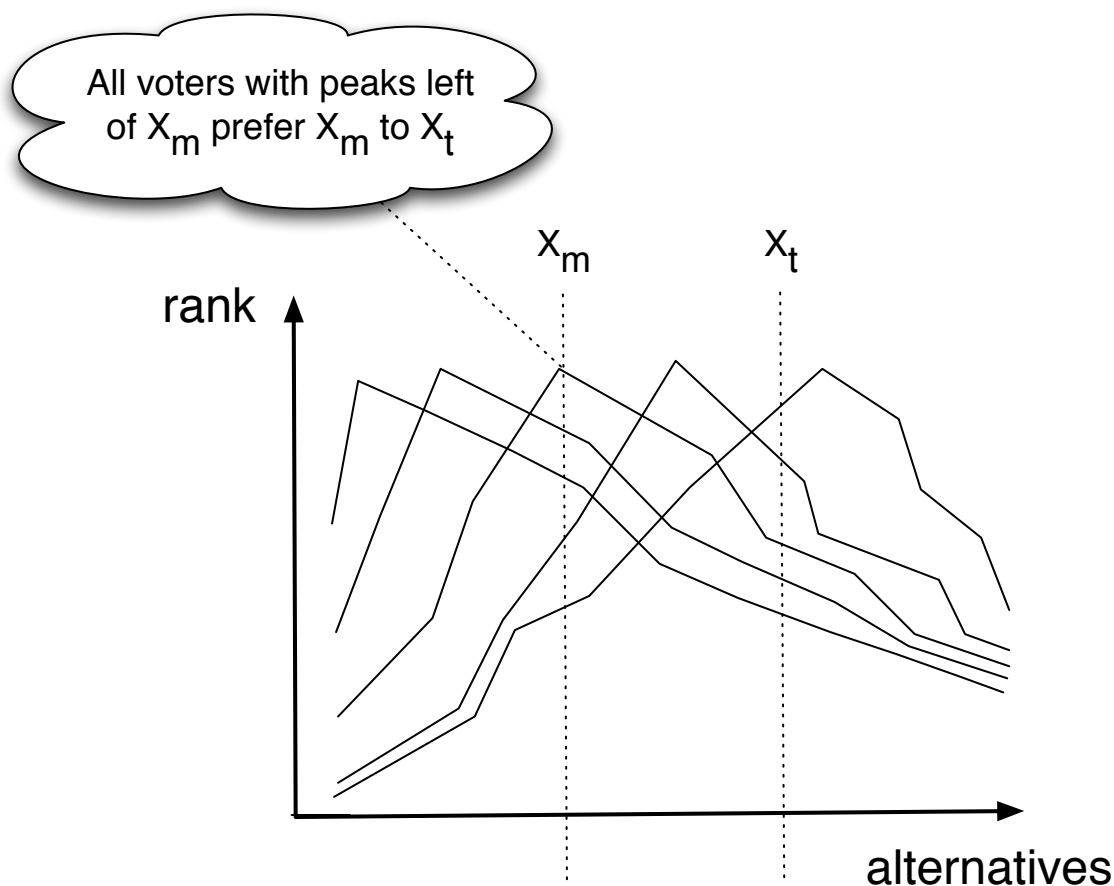


Figure 23.6: The proof that the median individual favorite  $X_m$  defeats every other alternative  $X_t$  in a pairwise majority vote: if  $X_t$  is to the right of  $X_m$ , then  $X_m$  is preferred by all voters whose peak is on  $X_m$  or to its left. (The symmetric argument applies when  $X_t$  is to the left of  $X_m$ .)

times in the sorted list: it is fine for the list to have repetitions. Now consider the individual favorite that forms the *median* of this list — that is, the individual favorite that lies exactly at the halfway point in the sorted order. For example, in the preferences from Figure 23.5, the sorted list of individual favorites would be  $X_1, X_2, X_3$ , and so the median is  $X_2$ . With more voters, if the individual favorites were (for example)  $X_1, X_1, X_2, X_2, X_3, X_4, X_5$ , then the median would also be  $X_2$ , since we are considering the median of the list with all the repetitions included.

The median individual favorite is a natural idea to consider as a potential group favorite, since it naturally “compromises” between more extreme individual favorites on either side. And in fact it works very well for this purpose:

*The Median Voter Theorem: With single-peaked rankings, the median individual*

*favorite defeats every other alternative in a pairwise majority vote.*

To see why this is true, let  $X_m$  be the median individual favorite, and let  $X_t$  be any other alternative. Let's suppose that  $X_t$  lies to the right of  $X_m$  — that is,  $t > m$ . (The case in which it lies to the left has a completely symmetric argument.) Let's also order the voters in the sorted order of their individual favorites.

The argument is now depicted schematically in Figure 23.6. The number of voters  $k$  is odd, and we know that — since it is the median —  $X_m$  is in position  $(k + 1)/2$  of the sorted list of individual favorites. This means that for everyone in the first  $(k + 1)/2$  positions,  $X_m$  is either their favorite, or their favorite lies to the left of  $X_m$ . For each voter in this latter group,  $X_m$  and  $X_t$  are both on the right-hand “down-slope” of this voter's preferences, but  $X_m$  is closer to the peak than  $X_t$  is, so  $X_m$  is preferred to  $X_t$ . It follows that everyone in the first  $(k + 1)/2$  positions prefers  $X_m$  to  $X_t$ . But this is a strict majority of the voters, and so  $X_m$  defeats  $X_t$  in a pairwise majority vote.

To put it succinctly: the median individual favorite  $X_m$  can always count on gathering a majority of support against any other alternative  $X_t$ , because for more than half the voters,  $X_m$  lies between  $X_t$  and each of their respective favorites.

From this fact about the median individual favorite, it is easy to see why majority rule among all pairs produces a complete and transitive group ranking: we simply build up the group ranking by identifying group favorites one at a time. That is, we start by finding the median individual favorite and placing it at the top of the group ranking. This is safe to do since the Median Voter Theorem guarantees that it defeats all other alternatives that will come later in the list. Now we remove this alternative from each individual ranking. Notice that when we do this, the rankings all remain single-peaked: essentially, we have simply “decapitated” the peak from each ranking, and the second item in each voter's ranking becomes their new peak. We now have a version of the same problem we faced before, with single-peaked rankings on a set of alternatives that is one smaller. So we find the median individual favorite on the remaining alternatives, place it second in the group ranking, and continue in this way until we exhaust the finite set of alternatives.

For example, applying this to the three voters in Figure 23.5, we would identify  $X_2$  as the median individual favorite, and we'd place it first in the group ranking. Once we remove this alternative, we have three single-peaked rankings on the alternatives  $X_1$ ,  $X_3$ ,  $X_4$ , and  $X_5$ . The individual favorites in this reduced set are  $X_1$ ,  $X_3$ , and  $X_3$ , so  $X_3$  is the new median individual favorite, and we place it second in the group ranking. Proceeding in this way, we end up with the group ranking

$$X_2 \succ X_3 \succ X_1 \succ X_4 \succ X_5.$$

Since voter 2 was the original “median voter” in the sense of having the original median individual favorite, the start of the group ranking necessarily agrees with the start of voter

2's individual ranking: they both place  $X_2$  first. However, the full group ranking does not coincide with voter 2's full individual ranking: for example, voters 1 and 3 both prefer  $X_1$  to  $X_4$ , even though voter 2 doesn't, and the group ranking reflects this.

## 23.7 Voting as a Form of Information Aggregation

Thus far, we have focused primarily on situations in which voting is used to aggregate fundamentally and genuinely different opinions within a group of people. But there are contexts in which voting is used by a group of people who have a shared goal — where it is reasonable to suppose that there is a true best ranking of the alternatives, and the purpose of the voting system is to discover it. This is unlikely to be appropriate for the ranking of political candidates or works of art, but it can be a good model of jury deliberations in cases where the decision hinges on genuine uncertainty about the facts. It can also be a good model for decisions made by a board of advisers to a company, evaluating business plans that each yield an uncertain future payoff.

In settings like these, where we imagine that there is a true best ranking, it's reasonable to suppose that the individual rankings differ only because they are based on different information, or based on different evaluations of the available information. If everyone in such a case had the same information and evaluated it in the same way, they would have the same ranking.

We will see that these considerations can lead to some potentially complex effects in the way individuals reason about their votes. As a starting baseline, we begin with a simple model in which individuals vote simultaneously, and based purely on their own individual rankings. We then discuss what happens in situations where one or both of these assumptions do not hold — when voting is done sequentially, or when knowledge of other rankings would cause an individual to change her own ranking.

**Simultaneous, Sincere Voting: The Condorcet Jury Theorem.** We begin with a simple setting in which there are two alternatives  $X$  and  $Y$ ; one of these two is genuinely the best choice, and each voter will cast a vote for what she believes to be this best choice.

To model the idea that voters possess different but uncertain information, we use a general framework that worked well in Chapter 16 on information cascades. We suppose first that there is a *prior probability* of  $X$  being the best choice, and that this is known to all voters. For simplicity, we'll take this prior probability to be  $1/2$  in our analysis here; this means that initially,  $X$  and  $Y$  are equally likely to be the best choice. Then, each voter receives an independent, private *signal* about which of  $X$  or  $Y$  is better. For some value  $q > 1/2$ , signals favoring the best choice occur at a rate of  $q$ ; writing this in terms of conditional probability

as in Chapter 16, we have

$$\Pr[X\text{-signal is observed} \mid X \text{ is best}] = q$$

and

$$\Pr[Y\text{-signal is observed} \mid Y \text{ is best}] = q.$$

(We can imagine each voter's signal as behaving like the flip of a biased coin: for each voter, it lands on the side indicating the better alternative with probability  $q$ .)

Unlike the case in Chapter 16, all the votes in our current analysis are being made simultaneously: no voter is able to see the decisions made by any other voter before reaching her own decision. Also, we assume that everyone is voting *sincerely*: each voter will choose the alternative she believes to be better, based on the information she has available (in the form of her private signal). To model sincere voting by an individual, we can use conditional probabilities just as in Chapter 16. When a voter observes a signal favoring  $X$ , she first evaluates the conditional probability

$$\Pr[X \text{ is best} \mid X\text{-signal is observed}].$$

She then decides to vote in favor of  $X$  if this probability is greater than  $1/2$ , and in favor of  $Y$  if this probability is less than  $1/2$ . The analogous reasoning applies if she observes a signal favoring  $Y$ ; we focus just on the case of an  $X$ -signal since the analysis is symmetric in the two cases.

We can evaluate the conditional probability underlying the voter's decision using Bayes' Rule, by strict analogy with the calculations we did in Section 16.3. We have

$$\Pr[X \text{ is best} \mid X\text{-signal is observed}] = \frac{\Pr[X \text{ is best}] \cdot \Pr[X\text{-signal is observed} \mid X \text{ is best}]}{\Pr[X\text{-signal is observed}]}.$$

By our assumption about the prior probability, we know that  $\Pr[X \text{ is best}] = 1/2$ . By the definition of the signals, we know that  $\Pr[X\text{-signal is observed} \mid X \text{ is best}] = q$ . Finally, there are two ways for an  $X$ -signal to be observed: if  $X$  is best, or if  $Y$  is best. Therefore,

$$\begin{aligned} \Pr[X\text{-signal is observed}] &= \Pr[X \text{ is best}] \cdot \Pr[X\text{-signal is observed} \mid X \text{ is best}] + \\ &\quad \Pr[Y \text{ is best}] \cdot \Pr[X\text{-signal is observed} \mid Y \text{ is best}] \\ &= \frac{1}{2} \cdot q + \frac{1}{2}(1 - q) = \frac{1}{2}. \end{aligned}$$

Putting all this together, we get

$$\Pr[X \text{ is best} \mid X\text{-signal is observed}] = \frac{(1/2)q}{1/2} = q.$$

The conclusion — which is completely natural — is that the voter will favor the alternative that is reinforced by the signal she receives. In fact, the calculation using Bayes' Rule gives

us more than just this conclusion; it also shows how much probability she should assign to this favored alternative based on the signal.

The Marquis de Condorcet wrote about this type of scenario in 1785. In his version, he took as a given the assumption that each voter chooses the best alternative with some probability  $q > 1/2$ , rather than deriving it from the assumption of a private signal — but the model based on either of these starting assumptions is effectively the same. Condorcet’s interest was in showing that majority rule is effective when there are many voters who favor the better of two choices at a rate slightly better than half. His probabilistic formulation of individuals’ decisions was a novel step — probability was still a relatively new idea in his time — and his main observation, now known as the *Condorcet Jury Theorem*, is the following. Suppose that  $X$  is the best alternative (the case for  $Y$  being symmetric). Then as the number of voters increases, the fraction of voters choosing  $X$  will converge almost surely to the probability of receiving an  $X$ -signal, which is  $q > 1/2$ . In particular, this means that the probability of the majority reaching a correct decision converges to 1 as the number of voters grows. In this sense, Condorcet’s Jury Theorem is one of the oldest explicit formulations of the “wisdom of crowds” idea: aggregating the estimates of many people can lead to a decision of higher quality than that of any individual expert.

## 23.8 Insincere Voting for Information Aggregation

One of the assumptions behind the Condorcet Jury Theorem in the previous section is that all individuals are voting sincerely: each is choosing the alternative he or she believes to be best, given the information available. On the surface, this seems like a mild assumption. If the voters could share all their signals, they would reach a unanimous evaluation of the best alternative; but since they aren’t able to communicate with each other and have access only to their own private signals, why should a voter do anything anything other than follow her best guess based on her own signal?

In fact, however, there are very natural situations in which an individual should actually choose to vote insincerely — favoring the alternative she believes to be worse — *even though her goal is to maximize the probability that the group as a whole selects the best alternative*. This is clearly a counter-intuitive claim, and its underlying basis has only been elucidated relatively recently [30, 159, 160]. To explain how this phenomenon can arise, we begin with a hypothetical experiment, modeled on a scenario described by Austen-Smith and Banks [30].

**An Experiment that Encourages Insincere Voting.** Here’s how the experiment works. An experimenter announces that an urn with 10 marbles will be placed at the front of a room; there is a 50% chance the urn will contain ten white marbles, and a 50% chance that it will contain nine green marbles and one white marble. (We describe the first kind of urn as



“pure” and the second kind as “mixed.”)

The experimenter asks a group of three people to collectively guess which kind of urn it is. Their group decision will be made according to the following protocol. First, each of the three people is allowed to draw one marble from the urn, look at it (without showing it to the other two), and then replace it in the urn. Then, the three people are asked to cast simultaneous votes, without communicating, each guessing which type of urn they think it is. If a majority of the votes are for the correct type of urn, then all three people win a monetary prize; if a majority of the votes are for the wrong type of urn, then all three people get nothing. (Note that each person gets nothing if the majority is wrong, even if they personally voted for the correct alternative.)

One can see that the experiment is designed to create a set of independent private signals for the voters: the color of the marble drawn by each voter is her private signal, and she cannot communicate this signal to any of the other voters. Rather, the group decision must be reached by majority vote, with each voter having access to these different and potentially conflicting probabilistic signals.

We now ask how an individual should reason about the conditional probabilities of the different urn types based on the marble she draws; after this, we’ll consider how she should actually vote.

**Conditional Probabilities and Decisions about Voting.** First, suppose that you (as one of the three people in the experiment) draw a white marble. While we won’t go through the precise calculations, it’s not hard to work out, using Bayes’ Rule just as we did in the previous section, that the urn in this case is significantly more likely to be pure than mixed. (Intuitively, if you see a white marble, it’s much more likely to have been from the all-white pure urn than to be the single white marble in the mixed urn.) On the other hand, if you draw a green marble, then in fact you know for certain that the urn is mixed — since green marbles are only found in mixed urns.

Therefore, if you were to vote sincerely, you would vote “pure” on drawing a white marble and “mixed” if you draw a green marble. But suppose you knew that the other two people in the group were going to vote sincerely, and you wanted to choose your vote to maximize the chance that the majority among the three of you produced the right answer. Then a useful question to ask yourself is, “In what situations does my vote actually affect the outcome?” If you think about it, your vote only affects the outcome when the other two (sincere) votes are split — when there is one vote for pure and one vote for mixed. In this case, however, one of your two partners in the experiments actually drew a green marble, and so the urn must be mixed. Here’s the conclusion from this reasoning: whenever your vote actually matters to the outcome, the urn is mixed!

So if you know your two partners will be voting sincerely, you can best help the group

by always voting “mixed,” so as to give a single draw of a green marble the chance to sway the majority outcome. In other words, you’re manipulating the group choice by voting strategically. You’re not doing this to take advantage of the other voters; indeed, you’re doing it to make it more likely the group will make the best choice. But nonetheless, it is not optimal for you to vote sincerely in this case, if your two partners are voting sincerely.

**Interpretations of the Voting Experiment.** Once we appreciate what’s going on here, it becomes natural to think of voting with a shared objective in terms of game theory. The voters correspond to players, their possible strategies are the possible ways of choosing votes based on private information, and they receive payoffs based on the votes selected by everyone. The experiment we’ve just considered constructs a scenario in which sincere voting is not an equilibrium. Notice that while our analysis has therefore ruled out the most natural candidate for an equilibrium, it hasn’t actually determined what an equilibrium looks like for this game. In fact, there are multiple equilibria, some of which are a bit complicated to compute, and we won’t try to work them out here.

There are a few further points worth reflecting on from this discussion. First, the experiment presents the phenomenon of insincere voting in a very clean and stylized form, which has the advantage of clearly exposing what’s going on. But versions of this scenario arise in real-world situations as well, when a highly symmetric decision process like majority vote clashes with a pair of alternatives that has an asymmetric structure — like the pure and mixed urns here. Suppose, for example, that a corporate advisory board has to decide between a risky and a safe course of action for the company, and they decide to use majority vote. Suppose further that board members have their own private evidence in favor of one option or the other, and in fact if anyone were to have genuine evidence in favor of the risky option, then it would be clearly the better choice. If you’re a board member in this case, and you know that the other board members will be voting sincerely, then your vote will only matter in the case when half of the rest of the board has evidence in favor of the risky option — in which case, the risky option is the better idea. So the group would be better served if you voted insincerely in favor of the risky option, to improve its chances of being chosen when it should be. Of course, viewing the process of voting as a game, you should appreciate that the situation is in fact more complicated: rather than assuming that the other board members will vote sincerely, you may want to assume that they are also going through this reasoning. Determining how to behave, given this, is a complex problem.

Finally, it’s worth highlighting a key methodological point in this analysis — the underlying principle in which you evaluate the consequences of your actions only in the cases where they actually affect the outcome. This was the clarifying insight that exposed why insincere voting was the right decision. Researchers have observed that the use of this principle for voting forms a parallel with reasoning in other game-theoretic contexts as well, including

the “winner’s curse” for auctions that we saw in Chapter 9 [159]. There, when many people bid on an item that has a common value (such as oil-drilling rights for a tract of land), the value of your bid only matters if you win, in which case your estimate of the true value of the item is more likely to be an over-estimate than an under-estimate. Hence you should take this into account when bidding, and bid lower than your estimate of the true value. This type of insincerity in bidding is analogous to the insincerity in voting that we’ve been discussing here; in both cases, they arise because you’re evaluating your decision contingent on its actually affecting the outcome, which provides additional implicit information that needs to be taken into account.

## 23.9 Jury Decisions and the Unanimity Rule

Jury decisions in criminal trials were an important initial example to motivate this discussion: they form a natural class of instances where a group of voters (the jurors) agree in principle that there is a “best” decision for the group — the defendant should be convicted if guilty and acquitted if innocent — and they want to aggregate their individual opinions to try arriving at this best decision. Given what we’ve just seen, it is natural to ask: can insincere voting arise in this case, and if so, what are its consequences? As Feddersen and Pesendorfer have argued, insincere voting in fact can arise naturally as a strategy for jurors who want their vote to contribute to the best overall group decision [160]. We describe the basic structure of their analysis here.

**Verdicts, Unanimity, and Private Signals.** If we compare jury decisions in criminal trials with the set-up for the Condorcet Jury Theorem from Section 23.7, we notice two basic differences, both of which arise from institutional features of the criminal-justice system designed to help avoid convicting innocent defendants.

The first difference is that it generally requires a unanimous vote in order to convict a defendant. So if we have  $k$  jurors, and the two options *acquittal* and *conviction*, each juror votes for one of these options, and conviction is chosen by the group only if each juror votes for it. The second difference is in the criterion that jurors are asked to use for evaluating the two alternatives. In the model from Section 23.7, if each voter could observe all the available information, she would choose alternative  $X$  if

$$\Pr[X \text{ is best} \mid \text{all available information}] > \frac{1}{2}.$$

In a criminal trial, however, the instructions to a jury are not, “The defendant should be convicted if he is more likely to be guilty than innocent,” but instead “The defendant should be convicted if he is guilty beyond a reasonable doubt.” This means that jurors should not

be asking whether

$$\Pr[\textit{defendant is guilty} \mid \textit{all available information}] > \frac{1}{2},$$

but whether

$$\Pr[\textit{defendant is guilty} \mid \textit{all available information}] > z$$

for some larger number  $z$ .

We now consider how to model the information available to each juror. Following the framework used for the Condorcet Jury Theorem in Section 23.7, we assume that each juror receives an independent private signal suggesting guilt (a  $G$ -signal) or innocence (an  $I$ -signal). The defendant, in reality, is of course either guilty or innocent, and we assume that signals favoring the truth are more abundant than signals favoring the wrong answer: for some number  $q > \frac{1}{2}$ , we have

$$\Pr[G\text{-signal} \mid \textit{defendant is guilty}] = q$$

and

$$\Pr[I\text{-signal} \mid \textit{defendant is innocent}] = q.$$

A juror who observes a  $G$ -signal is interested in the conditional probability of guilt given the signal, namely  $\Pr[\textit{defendant is guilty} \mid G\text{-signal}]$ . Let's assume a prior probability of  $1/2$  that the defendant is guilty — i.e., in the absence of any signals. Then the argument using Bayes' Rule from Section 23.7 (with conviction and acquittal playing the roles of the two alternatives  $X$  and  $Y$  from that section) applies directly here, showing that

$$\Pr[\textit{defendant is guilty} \mid G\text{-signal}] = q,$$

and similarly that

$$\Pr[\textit{defendant is innocent} \mid I\text{-signal}] = q.$$

The conclusions of the analysis to follow would remain essentially the same, with slightly different calculations, if we were to assume any prior probability between 0 and 1.

Before proceeding with the analysis, it's fair to ask whether the modeling assumption that jurors receive independent, private signals about guilt or innocence is reasonable — after all, they sit through the trial together, and they all see the same evidence being presented. Clearly, the assumption of private signals is a simplified approximation, but it is also clear that jurors in real trials can and do form widely divergent views of the facts in a case. This is natural: despite seeing the same evidence, jurors form different interpretations and inferences based on their own personal intuitions and decision-making styles — things that cannot necessarily be transmitted as facts from one person to another [160]. So in this case we can think of the private signals as representing private *interpretations* of the information presented, rather than some personal source of additional information. A rational juror is thus guided by her own signal, but she would also be influenced by knowledge of the signals of others — i.e. by knowledge that others had interpreted things the same or differently.

**Modeling a Juror's Decision.** As noted above, the unanimity rule is designed to make it hard for an innocent defendant to be convicted, since such a result would require every single juror to “erroneously” favor conviction. On the surface, this informal principle makes sense — but as we saw in Section 23.8, reasoning about such principles can become subtle when we assume that individuals are choosing their votes with the overall group decision in mind.

In particular, things become complicated for the following reason. Suppose that you're one of the  $k$  jurors, and you received an  $I$ -signal. At first, it seems clear that you should vote to acquit: after all, your  $I$ -signal on its own gives you a conditional probability of  $q > \frac{1}{2}$  that the defendant is innocent. But then you remember two things. First, the criterion for conviction by the group is

$$\Pr[\text{defendant is guilty} \mid \text{available information}] > z,$$

which means that in principle the unobserved signals of everyone else — if only you knew what they were — could be enough to push the conditional probability of guilt above  $z$ , despite your  $I$ -signal. Second, you ask yourself the key question from Section 23.8: “In what situations does my vote actually affect the outcome?” Given the unanimity rule, your vote only affects the outcome when every juror but you is voting to convict. If you believe that everyone else's vote will reflect the signal they received, then you can work out exactly what the full set of signals is in the event that your vote affects the outcome: it consists of  $k - 1$   $G$ -signals and your one  $I$ -signal.

What is the probability the defendant is guilty in this case? We can use Bayes' Rule to say that

$$\begin{aligned} & \Pr[\text{defendant is guilty} \mid \text{you have the only } I\text{-signal}] \\ &= \frac{\Pr[\text{defendant is guilty}] \cdot \Pr[\text{you have the only } I\text{-signal} \mid \text{defendant is guilty}]}{\Pr[\text{you have the only } I\text{-signal}]} \end{aligned}$$

Our assumption is that  $\Pr[\text{defendant is guilty}] = 1/2$ , and since the  $G$ -signals are independent, we have  $\Pr[\text{you have the only } I\text{-signal} \mid \text{defendant is guilty}] = q^{k-1}(1 - q)$ . (For this latter calculation, there is a probability of  $q^{k-1}$  that each of the  $k - 1$  other jurors gets a  $G$ -signal, times a probability of  $1 - q$  that you get an  $I$ -signal.) Finally, as usual in Bayes' Rule calculations, we determine the two different ways in which all jurors but you receive  $G$ -signals: if the defendant is guilty, or if he is innocent:

$$\begin{aligned} & \Pr[\text{you have the only } I\text{-signal}] \\ &= \Pr[\text{defendant is guilty}] \cdot \Pr[\text{you have the only } I\text{-signal} \mid \text{defendant is guilty}] + \\ & \quad \Pr[\text{defendant is innocent}] \cdot \Pr[\text{you have the only } I\text{-signal} \mid \text{defendant is innocent}] \\ &= \frac{1}{2} \cdot q^{k-1}(1 - q) + \frac{1}{2}(1 - q)^{k-1}q. \end{aligned}$$

(The second term in the last expression arises from an analogous calculation to what we used for the first term: if the defendant is innocent, there is a probability of  $(1 - q)^{k-1}$  that each of the  $k - 1$  jurors other than you gets a  $G$ -signal, times a probability of  $q$  that you get an  $I$ -signal.) Putting these quantities together, we have

$$\begin{aligned} \Pr[\text{defendant is guilty} \mid \text{you have the only } I\text{-signal}] &= \frac{\frac{1}{2}q^{k-1}(1 - q)}{\frac{1}{2}q^{k-1}(1 - q) + \frac{1}{2}(1 - q)^{k-1}q} \\ &= \frac{q^{k-2}}{q^{k-2} + (1 - q)^{k-2}}, \end{aligned}$$

where the second equality follows just by canceling  $q(1 - q)/2$  from both the numerator and denominator.

Now, since  $q > 1/2$ , the term  $(1 - q)^{k-2}$  represents an arbitrarily small portion of the total denominator as the jury size  $k$  goes to infinity — and so in particular

$$\Pr[\text{defendant is guilty} \mid \text{you have the only } I\text{-signal}]$$

converges to 1 as  $k$  goes to infinity. Hence if the jury size  $k$  is large enough, it follows that  $\Pr[\text{defendant is guilty} \mid \text{you have the only } I\text{-signal}] > z$ .

We conclude from this that if you believe everyone else is voting their signals, and if there are enough other jurors, then in the only case where your vote to acquit affects the outcome, the defendant is in fact guilty beyond a reasonable doubt. So if you were to vote with the actual instructions to the jury in mind, you should ignore your signal and vote to convict. Of course, you should do this with even more confidence in the event that you receive a  $G$ -signal, and so we can summarize the conclusion even more starkly: if you believe everyone else is voting their signals, and the jury size is large enough, you should always ignore your signal and vote to convict.

Intuitively, what's going on is that you only affect the outcome of a unanimous vote when everyone else holds the opposite opinion; on the assumption that everyone else is as well-informed as you are, and voting their true opinion, the conclusion is that they're probably (collectively) right, and you're wrong. As with our earlier example in Section 23.8, this serves as an interesting reminder that when you design a procedure or protocol for a group of people to follow, you should expect that they'll adapt their behavior in light of the rules you define. Here, the voting system based on unanimity was designed to help prevent erroneous convictions, but in fact it creates an incentive for people to disregard signals that the defendant is innocent.

**Equilibria for Voting under Unanimity and Other Systems.** As in Section 23.8, we've shown that (for large enough juries) voting your signal is not an equilibrium — if everyone else is doing it, then you should always vote to convict. In their analysis of this

problem, Feddersen and Pesendorfer went further and worked out what the equilibria for jury voting in this model actually look like.

First, there's an equilibrium that's easy to find but a bit pathological: if everyone decides to ignore their signals and vote to acquit, this is an equilibrium. To see why, notice that no juror can affect the outcome by changing her behavior, and hence there is no incentive for any juror to change what she is doing.

More interestingly, there is a unique equilibrium with the properties that (i) all jurors use the same strategy, and (ii) each juror's behavior actually depends on the signal she receives. This is a mixed-strategy equilibrium, in which each juror always votes to convict on a  $G$ -signal, and votes to convict with some probability between 0 and 1 on an  $I$ -signal. The idea is that each juror with an  $I$ -signal may randomly choose to disregard it, effectively correcting for the possibility that she is wrong. One can show that when jurors follow this equilibrium, the probability that their group decision convicts an innocent defendant is a positive number that does not converge to zero as the size of the jury goes to infinity. This forms a sharp contrast to the Condorcet Jury Theorem, where the probability of a correct decision is converging to 1 as the number of voters grows. The problem here is that the unanimity rule encourages voters to "over-correct" so strongly for the chance that they might be wrong, it leads to a noticeable probability that the group as a whole reaches the wrong decision.

Moreover, the unanimity rule is particularly bad in this regard. Specifically, with further analysis, we can study voting systems in which convicting a defendant requires only that an  $f$  fraction of the jurors vote for conviction, for different values of  $f$  with  $0 < f < 1$ . For a given choice of  $f$ , we'll call such a system the  $f$ -majority rule. There is still an equilibrium here in which jurors employ randomization, sometimes disregarding their signals to correct for the possibility that they are wrong. But with the  $f$ -majority rule, a juror's vote affects the outcome when the remaining jurors are divided between convicting and acquitting in a ratio of  $f$  to  $(1 - f)$  — a much less extreme split than under the unanimity rule, where a juror's vote affects the outcome only in the event that she is singular in her opposition to convicting. As a result of this, the randomized correction used by jurors is correspondingly less extreme, and one can show that as the jury size goes to infinity, the probability of the group decision being wrong goes to 0 [160].

This result offers a further reason to question the appropriateness of the unanimity rule — the result suggests that a decision rule for juries requiring conviction by a wide majority, rather than a unanimous vote, might actually induce behavior in which there is a lower probability of erroneous convictions. It is again an indication of the subtle issues that arise when one evaluates the trade-offs between different social institutions in light of the behaviors they induce in the people who take part in them.

## 23.10 Sequential Voting and the Relation to Information Cascades

Let's return to the original formulation of the Condorcet Jury Theorem, with individuals who vote simultaneously and sincerely over two alternatives  $X$  and  $Y$ . In the last two sections, we've examined what happens when we remove the assumption of sincerity. It's also interesting to instead remove the assumption of simultaneity and see what happens. We'll keep sincerity in this discussion, since this simplifies the analysis by only changing one aspect of the model at a time. So each voter will cast a vote for the alternative she believes to be the best choice.

When we assume that voters act sincerely but sequentially, we have a model that closely aligns with the formulation of information cascades from Chapter 16. In our model for information cascades, we assumed that voters make choices sequentially: they are able to observe the choices (but not the private signals) of earlier voters; and they can choose to disregard their own signals if it increases the chance that they personally choose the better alternative. Note that in this model of cascades, voters are still behaving sincerely in the sense that they are trying to choose the alternative that is more likely to be correct, based on everything they are able to observe.

Aside from this distinction between simultaneous and sequential voting, the set-up for the Condorcet Jury Theorem from Section 23.7 is otherwise quite similar to the model for information cascades from Chapter 16. In both models, there is a given prior probability for  $X$  to be correct, and there are private signals favoring the correct alternative with probability greater than  $1/2$ . Therefore, we can invoke our analysis from Section 16.5 to argue that if voters act sequentially, two initial votes in favor of  $X$  will cause a cascade in which all subsequent votes are for  $X$  as well — regardless of whether  $X$  is the correct decision. More generally, once the number of votes for one alternative first exceeds the number of votes for the other alternative by at least two, a cascade will form in which all subsequent voters will strategically choose to disregard their own signals.

The fact that cascades begin when one alternative leads the other by exactly two votes depends on the specific structure of our simplified model from Chapter 16. The broader principle, however, is quite general. In sequential voting of the type we're describing, a cascade will eventually develop. And cascades can be wrong: even if  $Y$  is the best alternative, a cascade for  $X$  can develop. Moreover, increasing the number of voters does essentially nothing to stop this cascade. So the principle behind the Condorcet Jury Theorem does not apply in this setting: There is no reason to expect that a large crowd of sequential voters will get the answer right.



Profile 1:

Individual	Ranking	Ranking restricted to $X$ and $Y$
1	$W \succ X \succ Y \succ Z$	$X \succ Y$
2	$W \succ Z \succ Y \succ X$	$Y \succ X$
3	$X \succ W \succ Z \succ Y$	$X \succ Y$

Profile 2:

Individual	Ranking	Ranking restricted to $X$ and $Y$
1	$X \succ Y \succ W \succ Z$	$X \succ Y$
2	$Z \succ Y \succ X \succ W$	$Y \succ X$
3	$W \succ X \succ Y \succ Z$	$X \succ Y$

Figure 23.7: The two profiles above involve quite different rankings, but for each individual, her ranking restricted to  $X$  and  $Y$  in the first profile is the same as her ranking restricted to  $X$  and  $Y$  in the second profile. If the voting system satisfies IIA, then it must produce the same ordering of  $X$  and  $Y$  in the group ranking for both profiles.

## 23.11 Advanced Material: A Proof of Arrow's Impossibility Theorem

In this section we give a proof of Arrow's Theorem [22, 23], which was stated in Section 23.5. The proof we present is not Arrow's original one; instead, we follow a shorter proof found more recently by John Geanakoplos [179].

Let's begin by stating the theorem in a language that will help in discussing the ideas in the proof. We start with a finite set of alternatives. We have a set of  $k$  individuals, whom we can assume to be numbered  $1, 2, 3, \dots, k$ ; each individual has a ranking of the possible alternatives. We'll call the collection of all  $k$  rankings a *profile*. In this terminology, a *voting system* is simply a function that takes a profile and produces a *group ranking*: a single ranking of the alternatives.<sup>1</sup> The voting system satisfies *Unanimity* if it puts  $X \succ Y$  in the group ranking whenever  $X \succ_i Y$  according to the ranking of each individual  $i$ . The voting system satisfies the *Independence of Irrelevant Alternatives* (abbreviated IIA) if the ordering of alternatives  $X$  and  $Y$  in the group ranking depends only on the ordering of  $X$  and  $Y$  in each individual ranking, and not on their position relative to any other alternatives.

Here's a slightly different but equivalent way to describe IIA, which will be useful in our discussion. Consider a profile of rankings, and any two alternatives  $X$  and  $Y$ . We say that an individual's ranking *restricted to  $X$  and  $Y$*  consists of a copy of his or her ranking in which we erase all the alternatives other than  $X$  and  $Y$ . A profile *restricted to  $X$  and  $Y$*  is

<sup>1</sup>As in earlier sections of this chapter, we will consider the case in which individual rankings have no ties, and the voting system is required to produce a group ranking that has no ties either.

the profile consisting of all individual rankings restricted to  $X$  and  $Y$ . Then, as illustrated in Figure 23.7, if a voting system satisfies IIA, it must produce the same ordering of  $X$  and  $Y$  for any two profiles that are the same when restricted to  $X$  and  $Y$ . (In other words, the profile restricted to  $X$  and  $Y$  is the only “data” the voting system can look at in ordering  $X$  and  $Y$  in the group ranking.)

Recall from Section 23.5 that a voting system can satisfy both Unanimity and IIA via *dictatorship*: it selects some individual  $j$  in advance, and for any profile of individual rankings, it simply declares the group ranking to be  $j$ ’s ranking. There are  $k$  different dictatorship procedures, depending on which of the  $k$  individuals is chosen in advance to be the dictator. Arrow’s Theorem is that the  $k$  dictatorship procedures are the only voting systems that satisfy Unanimity and IIA. This is the statement we prove here.

The challenge in proving Arrow’s Theorem is that the Unanimity and IIA conditions are both quite simple, and hence give us relatively little to work with. Despite this, we need to take an arbitrary voting system satisfying these two properties, and show that it in fact coincides with dictatorship by a single individual.

Our proof will consist of three main steps. First we show the following interesting fact; its utility in the proof is not immediately apparent, but it plays a crucial role. Let’s call  $X$  a *polarizing alternative* if it is ranked either first or last by every individual. Profiles  $P$  and  $P'$  in Figure 23.8 are examples of profiles in which  $X$  is a polarizing alternative. We’ll show that if a voting system satisfies Unanimity and IIA, then it must place any polarizing alternative in either first or last place in the group ranking. In other words, such a voting system can’t find a way to “average” and place a polarizing alternative somewhere in the middle of the group ranking. Note that many profiles don’t contain a polarizing alternative; this fact only applies to those that do. In the second step of the proof, we then use this fact to identify a natural candidate for the role of dictator, and in the third step, we prove that this individual is in fact a dictator.

**First Step: Polarizing Alternatives.** For the remainder of the proof, let  $F$  be a voting system satisfying Unanimity and IIA. We will use  $P$  to denote a profile of individual rankings, and use  $F(P)$  to denote the group ranking that  $F$  produces, as a function of this profile. We will work toward identifying an individual  $j$  with the property that  $F$  simply consists of dictatorship by  $j$ .

First, let  $P$  be a profile in which  $X$  is a polarizing alternative, and suppose by way of contradiction that  $F$  does not place  $X$  in either first or last place in the group ranking  $F(P)$ . This means that there are other alternatives  $Y$  and  $Z$  so that  $Y \succ X \succ Z$  in the group ranking  $F(P)$ .

Now, for any individual ranking that puts  $Y$  ahead of  $Z$ , let’s change it by sliding  $Z$  to the position just ahead of  $Y$ . This produces a new profile  $P'$ , as sketched in Figure 23.8. Since

Profile  $P$ :

Individual	Ranking
1	$X \succ \dots \succ Y \succ \dots \succ Z \succ \dots$
2	$X \succ \dots \succ Z \succ \dots \succ Y \succ \dots$
3	$\dots \succ Y \succ \dots \succ Z \succ \dots \succ X$

Profile  $P'$ :

Individual	Ranking
1	$X \succ \dots \succ Z \succ Y \succ \dots$
2	$X \succ \dots \succ Z \succ \dots \succ Y \succ \dots$
3	$\dots \succ Z \succ Y \succ \dots \succ X$

Figure 23.8: A polarizing alternative is one that appears at the beginning or end of every individual ranking. A voting system that satisfies IIA must put such an alternative at the beginning or end of the group ranking as well. The figure shows the key step in the proof of this fact, based on rearranging individual rankings while keeping the polarizing alternative in its original position.

$X$  is a polarizing alternative, the relative order of  $X$  and  $Z$  does not change in any individual ranking when we do this, nor does the relative order of  $X$  and  $Y$ . Therefore, by IIA, we still have  $Y \succ X \succ Z$  in the group ranking  $F(P')$ . But in  $P'$ , alternative  $Z$  is ahead of alternative  $Y$  in every individual ranking, and so by Unanimity we have  $Z \succ Y$  in the group ranking  $F(P')$ . Putting these together, the group ranking  $F(P')$  has  $Y \succ X \succ Z \succ Y$ , which contradicts the fact that the voting system  $F$  always produces a transitive group ranking.

This contradiction shows that our original assumption of alternatives  $Y$  and  $Z$  with  $Y \succ X \succ Z$  in  $F(P)$  cannot be correct, and so  $X$  must appear in either the first or last position in the group ranking  $F(P)$ .

**Second Step: Identifying a Potential Dictator.** In the next step, we create a sequence of profiles with the property that each differs from the next by very little, and we watch how the group ranking (according to  $F$ ) changes as we move through this sequence. As we track these changes, a natural candidate for the dictator will emerge.

Here is how the sequence of profiles is constructed. We pick one of the alternatives,  $X$ , and we start with any profile  $P_0$  that has  $X$  at the end of each individual ranking. Now, one individual ranking at a time, we move  $X$  from last place to first place while leaving all other parts of the rankings the same, as shown in Figure 23.9. This produces a sequence of rankings  $P_0, P_1, P_2, \dots, P_k$ , where  $P_i$

- (i) has  $X$  at the front of the individual rankings of  $1, 2, \dots, i$ ;

Profile  $P_0$ :

Individual	Ranking
1	$\dots \succ Y \succ \dots \succ Z \succ \dots \succ X$
2	$\dots \succ Z \succ \dots \succ Y \succ \dots \succ X$
3	$\dots \succ Y \succ \dots \succ Z \succ \dots \succ X$

Profile  $P_1$ :

Individual	Ranking
1	$X \succ \dots \succ Y \succ \dots \succ Z \succ \dots$
2	$\dots \succ Z \succ \dots \succ Y \succ \dots \succ X$
3	$\dots \succ Y \succ \dots \succ Z \succ \dots \succ X$

Profile  $P_2$ :

Individual	Ranking
1	$X \succ \dots \succ Y \succ \dots \succ Z \succ \dots$
2	$X \succ \dots \succ Z \succ \dots \succ Y \succ \dots$
3	$\dots \succ Y \succ \dots \succ Z \succ \dots \succ X$

Profile  $P_3$ :

Individual	Ranking
1	$X \succ \dots \succ Y \succ \dots \succ Z \succ \dots$
2	$X \succ \dots \succ Z \succ \dots \succ Y \succ \dots$
3	$X \succ \dots \succ Y \succ \dots \succ Z \succ \dots$

Figure 23.9: To find a potential dictator, one can study how a voting system behaves when we start with an alternative at the end of each individual ranking, and then gradually (one person at a time) move it to the front of people's rankings.

(ii) has  $X$  at the end of the individual rankings of  $i + 1, i + 2, \dots, k$ ; and

(iii) has the same order as  $P_0$  on all other alternatives.

So in other words,  $P_{i-1}$  and  $P_i$  differ only in that individual  $i$  ranks  $X$  last in  $P_{i-1}$ , and he ranks it first in  $P_i$ .

Now, by Unanimity,  $X$  must be last in the group ranking  $F(P_0)$ , and it must be first in the group ranking  $F(P_k)$ . So somewhere along this sequence there is a first profile in which  $X$  is not in last place in the group ranking; suppose this first profile is  $P_j$ . Since  $X$  is a polarizing alternative in  $P_j$ , and it is not in last place in  $F(P_j)$ , it must be in first place.

So individual  $j$  has a huge amount of power over the outcome for alternative  $X$ , at least in this sequence of rankings: by switching her own ranking of  $X$  from last to first, she causes

$X$  to move from last to first in the group ranking. In the final step of the proof, we will show that  $j$  is in fact a dictator.

**Third Step: Establishing that  $j$  is a Dictator.** The key argument in showing that  $j$  is a dictator is to show that for any profile  $Q$ , and any alternatives  $Y$  and  $Z$  that are different from  $X$ , the ordering of  $Y$  and  $Z$  in the group ranking  $F(Q)$  is the same as the ordering of  $Y$  and  $Z$  in  $j$ 's individual ranking in  $Q$ . After that we'll show that the same also holds for pairs of alternatives in which one of the alternatives is  $X$ . In this way, we'll have established that the ordering of each pair is determined entirely by  $j$ 's ordering, and hence  $j$  is a dictator.

So let  $Q$  be any profile, and let  $Y$  and  $Z$  be alternatives not equal to  $X$  such that  $j$  ranks  $Y$  ahead of  $Z$ . We will show that  $F(Q)$  puts  $Y$  ahead of  $Z$  as well.

We create an additional profile  $Q'$  that is a variant of  $Q$ ; this new profile will help us understand how  $j$  controls the ordering of  $Y$  and  $Z$ . First, we take  $Q$ , move  $X$  to the front of the individual rankings of  $1, 2, \dots, j$ , and move  $X$  to the end of the individual rankings of  $j+1, j+2, \dots, k$ . Then, we move  $Y$  to the front of  $j$ 's individual ranking (just ahead of  $X$ ). We call the resulting profile  $Q'$ .

Now, we make the following observations.

- We know that  $X$  comes first in the group ranking  $F(P_j)$ . Since  $Q'$  and  $P_j$  are the same when restricted to  $X$  and  $Z$ , it follows from Independence of Irrelevant Alternatives that  $X \succ Z$  in  $F(Q')$ .
- We know that  $X$  comes last in the group ranking  $F(P_{j-1})$ . Since  $Q'$  and  $P_{j-1}$  are the same when restricted to  $X$  and  $Y$ , it follows from IIA that  $Y \succ X$  in  $F(Q')$ .
- By transitivity, we conclude that  $Y \succ Z$  in  $F(Q')$ .
- $Q$  and  $Q'$  are the same when restricted to  $Y$  and  $Z$ , since we produced  $Q'$  from  $Q$  without ever swapping the order of  $Y$  and  $Z$  in any individual ranking. By IIA, it follows that  $Y \succ Z$  in  $F(Q)$ .
- Since  $Q$  was any profile, and  $Y$  and  $Z$  were any alternatives (other than  $X$ ) subject only to the condition that  $j$  ranks  $Y$  ahead of  $Z$ , it follows that the ordering of  $Y$  and  $Z$  in the group ranking is always the same as  $j$ 's.

Thus we've shown that  $j$  is a dictator over all pairs that do not involve  $X$ . We're almost done; we just have to show that  $j$  is also a dictator over all pairs involving  $X$  as well.

To show this, first observe that we can run the argument thus far with respect to any other alternative  $W$  different from  $X$ , and thereby establish that there is also an individual  $\ell$  who is a dictator over all pairs not involving  $W$ . Suppose that  $\ell$  is not equal to  $j$ . Now, for  $X$  and some third alternative  $Y$  different from  $X$  and  $W$ , we know that the profiles  $P_{j-1}$

and  $P_j$  differ only in  $j$ 's individual ranking, yet the ordering of  $X$  and  $Y$  is different between the group rankings  $F(P_{j-1})$  and  $F(P_j)$ . In one of these two group rankings, the ordering of  $X$  and  $Y$  must therefore differ from the ordering of  $X$  and  $Y$  in  $\ell$ 's individual ranking, contradicting the fact that  $\ell$  is a dictator for the pair  $X$  and  $Y$ . Hence our assumption that  $\ell$  is different from  $j$  must be false, and thus  $j$  is in fact a dictator over all pairs.

## 23.12 Exercises

1. In this chapter, we discussed how voting systems based on majority rule are susceptible to strategic agenda-setting. Let's explore how one might do this on some basic examples.

- (a) Suppose there are four alternatives, named  $A$ ,  $B$ ,  $C$ , and  $D$ . There are three voters who have the following individual rankings:

$$B \succ_1 C \succ_1 D \succ_1 A$$

$$C \succ_2 D \succ_2 A \succ_2 B$$

$$D \succ_3 A \succ_3 B \succ_3 C$$

You're in charge of designing an agenda for considering the alternatives in pairs and eliminating them using majority vote, via an elimination tournament in the style of the examples shown in Figure 23.3.

You would like alternative  $A$  to win. Can you design an agenda (i.e. an elimination tournament) in which  $A$  wins? If so, describe how you would structure it; if not, explain why it is not possible.

- (b) Now, consider the same question, but for a slightly different set of individual rankings in which the last two positions in voter 3's ranking have been swapped. That is, we have:

$$B \succ_1 C \succ_1 D \succ_1 A$$

$$C \succ_2 D \succ_2 A \succ_2 B$$

$$D \succ_3 A \succ_3 C \succ_3 B$$

We now ask the same question: Can you design an agenda in which  $A$  wins? If so, describe how you would structure it; if not, explain why it is not possible.

2. The Borda Count is susceptible to strategic misreporting of preferences. Here are some examples to practice how this works.