1

Moving Out of Flatland

Be patient, for the world is broad and wide.
– *The Square*

On May 1, 2011, Keith Urbahn, chief of staff of Donald Rumsfeld, wrote 77 characters announcing to the world that a turning point in contemporary history had been reached (Nahon and Hemsley, 2013):

> So I'm told by a reputable person they have killed Osama Bin Laden. Hot damn.

These 77 characters started a chain reaction that led, within minutes, to the worldwide diffusion of the news and marked the beginning of the post-Osama era. First posted on Twitter, the news quickly reached millions of other Twitter users, then spread over tens of other online social networks, appeared in traditional media (e.g., television, radio), and became a common topic of discussion in the offline world, both in its original 77-character format and rephrased so that only its information content was preserved.

We can think of this event from many perspectives. We could focus on its historical value or observe how social media, like Twitter, are challenging the traditional relationship between politics and journalism. We could also use this tweet and the reactions to it to describe how information virality works in contemporary society and why the Internet has made it different from everything else we have seen in the history of humanity. For the concerns of this book, we can say that Mr. Urbahn provided yet another example of the multidimensional nature of our social experience.

1.1 Multiple Social Networks in Our Everyday Experience

Our social experience is inherently a multifaceted reality made of multiple interconnected networks defining our understanding of the world and our role in it. These networks do not exist autonomously; they are defined by our social relations and connected into a larger system by our activities. This is exactly what Keith Urbahn did when he tweeted something that a reputable person had told him: he defined a bridge between two networks. More precisely, he moved a specific piece of information out of an offline and exclusive network into a worldwide online digital network, and in switching between them, he was surely aware of the consequences. Dealing with multiple social networks is part of our daily experience: we continuously and effortlessly juggle our networks; we bridge them to move valuable information from one network to another; or we keep them separate to protect our privacy, to preserve face, or to offer a specific representation of ourselves to our potential audiences (Hogan, 2010). However, the fact that we continuously deal with multiple social networks and that we do it with no effort does not mean that this is a trivial activity or that we can overlook it. Quite the opposite is true: our *networked society* – described by Rainie and Wellman (2012) as being characterized by multiple overlapping social and media networks – provides us with a number of examples why these networks should be more and more relevant.

Let us be clear that the coexistence of multiple social networks is not a discovery of this book nor the result of recent research efforts. For example, the fact that we are connected to other people through multiple types of relational ties, although representing only one possible view of the problem, was known long before the field of social network analysis (SNA) was developed and has always been acknowledged in the SNA literature as a foundational feature of the discipline. Quoting Wasserman and Faust (1994, p. 18),

> the range and type of ties can be quite extensive, [including] evaluation of one person by another[,] transfers of material resources[,] association or affiliation[, etc.].

However, when we move from the qualitative description of the discipline to its mathematical definitions, quantitative measures, and practical applications, traditional SNA has typically focused on one type of actor and tie at a time:

> Most social network applications focus on collections of actors that are all of the same type.…With multirelational data, we suggest that [actor centrality and prestige] be calculated for *each* relation. (p. 17)

Although this sort of structural approach to social understanding is still a very powerful tool for unfolding the hidden structures behind our social activities, over recent years, it has become more and more evident how a monodimensional analysis is unable to account for a growing number of phenomena.

Looking only at a single type of relational tie within a single social network risks either defining a world where different kinds of relationships are ontologically equivalent or overlooking the invisible relationships emerging from the interactions among different types of ties. This apparently harmless simplification can alter the topology of the network, producing inaccurate or misleading results (Magnani and Rossi, 2011). For example, Mr. Urbahn would not be regarded as an important user on Twitter if we were to measure importance solely by counting his number of followers, but in the larger system, including both his Twitter and offline social relations, he played a key role as a bridge between two complementary networks, one providing trust and the other speed of diffusion. Without an expanded perspective, we would not be able to describe the whole range of problems and structures that can be found in a world of multiple social networks. We would not even be able to conceive of multidimensional ideas within a monodimensional space. This is exactly the situation described in Abbott's (1884) famous novella, where a square explains how our senses and conceptual tools define what we can comprehend:

> I admit the truth of your critic's facts, but I deny his conclusions. It is true that we have really in Flatland a Third unrecognized Dimension called "height," just as it is also true that you have really in Spaceland a Fourth unrecognized Dimension, called by no name at present, but which I will call "extra-height." But we can no more take cognizance of our "height" than you can of your "extra-height." Even I who have been in Spaceland, and have had the privilege of understanding for twenty-four hours the meaning of "height"; even I cannot now comprehend it, nor realize it by the sense of sight or by any process of reason; I can but apprehend it by faith. (p. 7)

The good news is that we are luckier than our square friend: a gestalt SNA is possible, because our social experience is indeed multidimensional. Similarly to the Spacelanders who were looking for thickness on the two-dimensional objects in Flatland, we have always had the perception that there has to be something more. We have been confined by the world we created ourselves, but we have always had the clear understanding that a single dimension is not enough, as clearly stated in foundational SNA sources. This perception has been different in different disciplines: physical sciences, accustomed to looking for unifying models, have sometimes regarded any dyadic phenomena as a network, making no distinction between friends interacting on Facebook and proteins interacting in a cell; social scientists have struggled with this for many years in trying to understand how different kinds of ties affect each other (Borgatti et al., 2009). Nevertheless, for a long time, these interactions have largely been studied within what we call a single-layer perspective.

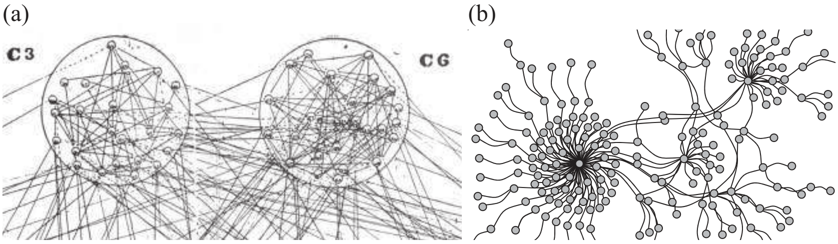

Figure 1.1. Visualizations of an offline and an online social networks: (a) a sociometric diagram and (b) a Twitter reply network (2012). (Part (a) reproduced with permission of the American Society of Group Psychotherapy and Psychodrama from *Who Shall Survive?*, J. L. Moreno, M.D. Beacon House Inc. Beacon, N.Y., Second Edition, 1953).

For almost a century, one of the most effective SNA tools to measure our social interactions has been the *simple graph*, where *simple* is a mathematical term and does not imply that there is anything simple at all in our social lives. A simple graph is defined as a set of *nodes* (representing individuals or organizations, often called *actors* in the SNA tradition) with *edges* between them, also called *links* or *connections* (representing *relational ties*, e.g., friendship relationships) and with no edges connecting a node to itself. To provide some historical perspective, Figure 1.1 shows both one of the first known examples of a graph-based sociometric diagram, hand-drawn by Moreno (1934), and a more recent Twitter reply network drawn using one of the many currently available graph visualization tools[1] (Rossi and Magnani, 2012). Despite their mundane nature, it would be extremely complex (if not impossible) to accurately describe the aforementioned tweet about Bin Laden's death and its related events using a simple graph. How could we represent the differences between Mr. Urbahn's reputable source and his Twitter followers? Would it just be a matter of weights? And how could we represent the differences in network structure, localized social practices, and technological affordances that are necessary to fully explain what happened? These are all questions that cannot find a complete answer inside Flatland.

Moving out of Flatland does not mean that Flatland is wrong. It does not mean that we cannot explain anything within its boundaries: traditional SNA has repeatedly proved itself to have great explanatory power. To some extent, it is not even a matter of avoiding potential misleading results – even if that would be an indirect benefit. It is more a matter of introducing a new perspective, new ideas, and new dimensions that were not possible before. This is why this book should be perceived more as an extension of traditional SNA and network science into new directions than their as about their evolution.

[1] The igraph network analysis package, available at: http://www.igraph.org/.

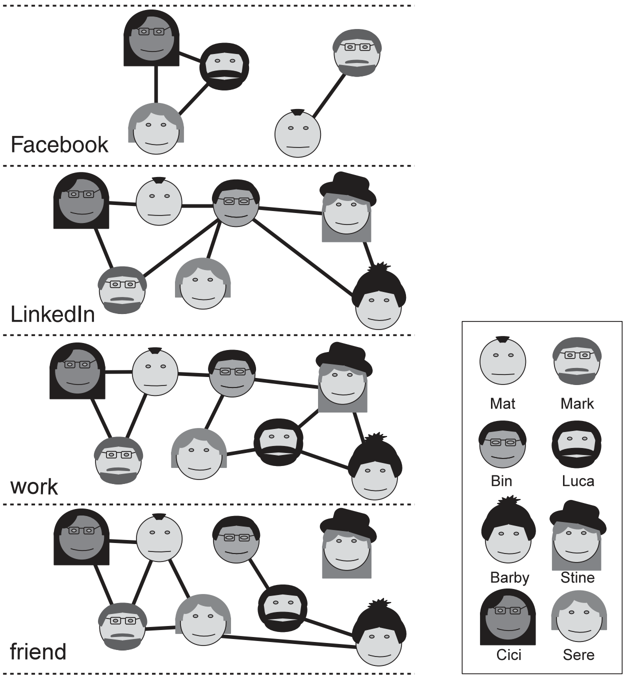

Figure 1.2. A multilayer social network with four layers and eight actors, used as a running example throughout the book.

1.2 An Introductory Example

A quick look at a simple example will help us provide a brief overview of the kinds of analyses enabled by what we are going to call the *multilayer social network model*. In Chapter 2, we precisely define what we mean by multilayer social network, but for now, we can think of it as a social network with nodes and/or edges organized into multiple *layers*, where each layer represents a different kind of node or edge, a different social context, a different community, a different online social network (OSN), and so on.

Consider the four layers in Figure 1.2, representing two offline relations (work collaborations and friendships) and two online, social media – based layers (LinkedIn and Facebook). These four layers are not independent, but they are connected through the common actors indicated in the right-hand side box. Each node in each of the four layers represents one of these actors or a social media account owned by one of them. These actors define bridges between the four layers, in the same way as Mr. Urbahn defined a bridge between the offline network of his reputable source and his Twitter followers.

Looking at the figure, we can observe some interesting patterns emerging from the dependencies between the four layers. For example, Cici, Mat, and Mark seem to form a cohesive group spanning multiple layers: they are all connected to each other (i.e., they form a *clique*) on the work and friend layers, and Cici is connected to both Mark and Mat also on LinkedIn. We can say that they form a strong group spanning several relational dimensions. However, Cici does not use Facebook to interact with Mark and Mat: different social networks can be strategically used in different ways, for example, to reach different audiences or to prevent some people from accessing information produced on a specific layer. If we count the total number of ties for each actor, we can see that both Sere and Cici have eight of them over the four layers. However, Cici is only connected to four actors (Mark, Mat, Luca, and Sere), whereas Sere is directly connected to almost everyone. In particular, she is directly connected with the same number of actors as Bin, despite that Bin has five connections just on the LinkedIn layer, more than anyone else – in contrast, Sere has at most three contacts on each single layer. At the same time, Sere can easily spread information to everyone in the four layers, assuming that this information is appropriate for all of them, for example, gossip might spread on Facebook more easily than on LinkedIn. Finally, we can look at the distance between Cici and Stine. Interestingly, without changing layers, they can reach each other only on the LinkedIn and work layers, through Mat and Bin. However, Cici can use her Facebook relationship with Luca to reach Stine in that context. The existence of multidimensional chains of social relationships (e.g., Cici is a Facebook friend with Luca, who works with Stine) is at the basis of many recent advances in this area, for example, identifying Luca's role in connecting people from different layers.

Although this multilayer network is too small to allow us to reach any significant conclusions on general network properties, this example shows how the joint analysis of multiple layers can provide knowledge that is not present in each layer when layers are considered independently of each other.

In addition, multilayer networks can significantly affect our understanding of a social system even when only single-layer data are available. A clear example are Facebook friendship connections, which may indicate friends, acquaintances, colleagues, family members, and so on. As a consequence, performing tasks like identifying communities becomes very complex because of the many overlapping social contexts, and if some data are missing only from one specific hidden layer, then some descriptive network measures can be either under- or overestimated, depending on the layer where the data are missing. In summary, advances in multilayer network analysis are also leading to a rethinking of how we analyze single-layer networks.

We now hope that the reader is looking forward to reading the rest of the book and getting more details about how to deal with these kinds of data. However, we first must raise one question. So far we have claimed that the topic of this book is very important, we have suggested that it may extend traditional SNA with a number of results not achievable otherwise, and we have mentioned how its importance has been acknowledged for a long time. So, why are we writing this book now? Why was it not written ten, twenty, or thirty years ago?

One possible answer to this question is related to the invisible nature of relational ties. Sociologists have known for a long time that invisible connections, often hard to describe, lie behind many social phenomena (Wellman and Berkowitz, 1988). SNA relies on making relational ties visible so that they can be studied as a social graph. However, recording these ties is complicated: until approximately a decade ago, network data collection was typically performed using questionnaires. Getting accurate information about relational ties, even a single type, would take great effort, and with scarce data, it is often difficult to develop popular models and analysis methods. Today, the explosion of OSNs and Web 2.0 has revealed the existing networks of relations bonding societies together and has highlighted their multidimensional nature (Rainie and Wellman, 2012). Although most of these relations have not been created by digital technologies, digital technologies have nevertheless made them visible, popularizing the concept of the network as a meaningful way to think about our social experiences. We all can see our networked world: links are everywhere, all around us, as Barabási (2002) pointed out in his popular book. So, it has become easier to reason about the relationships between all these links, especially when they belong to different social networks. In addition, the availability of large non–social network data sets, such as interconnected traffic networks (airplanes, trains, cars, etc.) and biological networks, has boosted the development of general measures and methods that have the potential to be applied also to a social context.

1.3 Scope and Other Learning Resources

The objective of this book is to provide an accessible presentation of recent research results on the analysis and mining of multilayer social networks. By *accessible*, we do not mean that we simplify the available material, but we try to provide a presentation that does not require a specific background to be appreciated or understood. Different disciplines, such as sociology, computer science, and physics, have contributed to this area, the first providing semantics and interpretative keys to new SNA measures, the second introducing new data mining algorithms, and the third formalizing general models and global dynamics of complex network systems. Interdisciplinary contributions have also been increasingly frequent in recent years. With this book, we want to take a step forward toward developing a homogeneous and interdisciplinary body of work on multilayer social networks. As such, although we try to be exhaustive whenever possible – something probably not completely achievable, given the vitality of the area – our main objective is to provide a consolidated presentation of the available material that makes sense from multiple points of view.

Some of the methods for multilayer social networks presented in this work can be applied to other kinds of (nonsocial) networks, making sure to take care to rethink their semantics so that they fit the different domains, and methods for generic multilayer networks can also be used to describe social networks. The reader interested in a more dense, general, and theoretical presentation of multilayer networks with less focus on data mining and social interpretations can refer to the excellent survey papers by Kivelä et al. (2014) and Boccaletti et al. (2014). Another valuable resource covering some of the literature on mining a specific kind of multilayer network called the *heterogeneous information network* has been published by Sun and Han (2012). For the reader more interested in concrete kinds of (nonsocial) networks and practical applications, a valuable collection of papers has been edited by D'Agostino and Scala (2014) that focuses on another specific kind of multilayer network called the *network of networks*. The confused reader who is getting lost with all these different names of related models may first check the next chapter of this book, where we provide more details about their differences and similarities. In addition, recent literature surveys have focused on the specific aspects of information diffusion (Salehi et al., 2015) and community detection (Bothorel et al., 2015) in multilayer networks. For more specific references, we refer the reader to the literature discussed in the different chapters.

As a final note, although we have tried to assemble an interdisciplinary group of authors, our presentation of the related material is biased toward our own backgrounds and certainly misses some important references of which we are not aware. This is a risk that must be run in an attempt to provide a uniform presentation of material developed in many different fields, and we apologize in advance for underrepresenting some disciplines or areas.

1.4 Outline of the Book

The book is divided into four parts. Part I ("Models and Measures") describes how to represent and compute quantitive descriptions of multilayer social networks. Part II ("Mining Multilayer Networks") explains how to discover hidden patterns such as communities or associations between edges on different layers. Part III ("Dynamical Processes") presents models of how multilayer social networks coevolve in time and how information, ideas, and behaviors diffuse in them. Finally, Part IV ("Conclusion") discusses our personal view on the future evolution of the discipline.

We have tried to keep each chapter as self-contained as possible. However, Section 2.1 introduces the models and terminology used in the book and may be useful to read first. Throughout the book, we refer to concepts defined in Chapter 3 ("Measuring Multilayer Social Networks"). Therefore, we have organized them so that they can be easily identified from the table of contents and checked without reading the full chapter, as needed. Although we will be happy if the reader decides to read the whole book, we will be even happier if he or she can save some time to use for something else, for example, reading some poetry or listening to some good music.

We conclude this introduction by providing additional details about the content of each chapter. Part I reviews alternative ways to represent and measure multilayer networks. In Chapter 2, we present the terminology and data model used in the book, in addition to various other data models for social networks allowing multiple types of nodes (also called heterogeneous, attributed, or multitype networks), multiple types of relational ties (also called multiplex or multidimensional networks), or explicitly representing the coexistence of separate, interdependent social networks. All these can be seen as specific cases of the general multilayer model used in this book. While going through all these different models, this chapter provides a historical account of the different approaches developed to study multilayer social networks in different disciplines. We also describe several application areas and provide pointers to the main existing data sets, some of which are used as working examples throughout the book. Chapter 3 presents the main measures for the quantitative description of multilayer social networks, complementing and extending traditional SNA metrics. In that chapter, we define and exemplify degree and neighborhood centrality, multidimensional distances, and derived measures such as betweenness, transitivity, relevance, and layer correlation. Other measures used to identify communities (modularity) and to predict the creation of new edges are treated in the corresponding chapters in Part II.

Part II focuses on identifying hidden patterns in multilayer social networks. We start focusing on the important aspects of data collection, preprocessing, and exploration. Chapter 4 discusses issues in data collection related to using sampling and to the presence of missing data and different approaches to transforming the collected data. Analyzing multiple layers is inherently more complex than dealing with a single layer, and too much information can generate noise and hide some important patterns that are present only in some of the layers or in their combinations. Therefore, it can be useful to simplify the data, from the extreme choice of creating a single *flattened* social graph to more sophisticated data transformations that remove or merge only some of the layers or portions of them. Then, after the data have been collected and prepared, a typical way to explore them is through visualization. Although only a few visualization methods have been specifically defined for multilayer networks, as reported in Chapter 5, they can be valuable in highlighting some basic patterns in the data, for example, the presence of well-separated communities or strong correlations or differences among relational ties in different layers. Chapters 6 and 7 focus on popular data mining tasks: identifying communities, predicting future relational ties, and computating layer correlations. Although community detection is one of the most widely studied network mining problems and has undoubtedly achieved many important results over the years, the complex conceptual and methodological problems associated with community detection methods rise to a new level when we consider multilayer networks. In Chapter 6, we provide wide coverage of existing approaches. Apart from community detection, other data mining problems have received less attention so far but are likely to become popular in the near future. In Chapter 7, we focus on predicting the appearance of new relational ties (a data mining problem known as *link prediction*) and discovering associations between ties in different layers.

Part III explores dynamical processes on multilayer social networks. Network formation models are among the most important tools in the field known as network science. A typical application of artificially generated networks is to provide *null* models that can be used to test new measures and make comparisons with real networks so that significant patterns can be highlighted in the real data. In addition, network growing models are useful for experimenting on the dynamics underlying the evolution of social relationships. In Chapter 8, we review some recent works modeling the coevolution of multiple layers representing interdependent social networks. Another important type of process happening in social networks is the diffusion of information, which can reach a very high speed when online social networks are used. However, information is not the only spreadable entity: diffusion processes in social networks can also involve opinions and behaviors. Traditionally, diffusion has been modeled with percolation or epidemic models, both of which have been shown to exhibit novel phenomena on multilayer networks, including new types of phase transitions and entirely new phases. In Chapter 9, we present material in these areas and explain how it can be applied to information and opinion diffusion.