

Summary

The abundance of data has changed the way we conduct our personal and social life. In fact, in recent decades, we have faced an explosion of *social data*, in which unprecedented variety of personal information has become accessible to the public. Social data consists of two different data categories: First is the social data on individuals. This category captures the history of their actions, their habits, and their future plans. The second category is the social data on interactions among individuals, such as friendship interactions, professional connections and co-location observations. Once we integrate these two categories of social data, *social graphs* emerge. The importance of taking control over integration, modeling and manipulation of the social graphs has attracted interest from various research communities. As a result, much progress has been made in the analysis of social graphs. However, a faster progress has been made in data collection technologies resulting in a higher volume of data, more variety in forms of data and less veracity in the available sources.

With the rise of online social networks and widely accessible historical archives, the complexity and dynamics of social graphs have led to the following two basic research challenges. First, how to extract the structure and dynamics of social graphs from heterogeneous social data, that is, how to turn social data into social graphs. Second, how to model and analyze the properties of dynamic social graphs, that is, how to analyze evolving social graphs.

The analysis of dynamic social graphs is a common theme in this thesis. Chapter 2 provides the background knowledge on information retrieval, social graphs, and the theory of dynamical systems that are required for a better understanding of this thesis. It also introduces the datasets that are used throughout the thesis for experimental analysis and numerical verifications. The data-driven approaches proposed in Chapters 3 and 4 provide powerful practical tools for the integration of heterogeneous sources of social data. These tools resulted in the automatic generation of a set of rich social graphs that could not be analyzed with existing analytical methods.

Two novel analytical models are proposed in Chapter 5 that allow for a deeper understanding of the topological properties of social graphs. A new analytical model is proposed in Chapter 6 to model the evolution of behaviors in such graphs. The proposed models successfully capture important properties of dynamic social graphs. A further research step is taken in Chapter 7: the simultaneous evolution of topology and behaviors in social graphs is studied and the effect of this co-evolution on topological diversity of the social graphs and their equilibrium state is demonstrated.

To conclude, in this thesis, effective methods for turning heterogenous social data into social graphs are developed and novel models for analysis of such dynamic social graphs are proposed. This thesis opens up at least four new promising research avenues. First, it allows for the conduction of *Advanced Genealogical Research*. Second, it paves the way to *Trace the Origins of Heavy-Tailed Distributions* in social networks using theory of dynamical systems. Third, it provides new methods for studying *Controllability of Behaviors in Social Networks*. Fourth, it establishes a bridge between *Swarm Robotics* and *Experimental Economics*.

Samenvatting

De overvloed aan data heeft ons persoonlijke en sociale leven veranderd. In de afgelopen decennia zijn we geconfronteerd met een explosie van *sociale data*, waarin een ongekeerde hoeveelheid aan persoonlijke informatie publiekelijk toegankelijk is geworden. Sociale data bestaan uit twee verschillende categorieën. De eerste is sociale data over individuen. Deze categorie bevat hun gedrag, hun gewoonten en hun toekomstplannen. De tweede categorie beschrijft de interactie tussen individuen, zoals vriendschap, professionele relaties en ontmoetingen. Wanneer we deze twee categorieën van sociale gegevens samenvoegen ontstaan *sociale grafen*. Het belang van het controleren van de integratie, modellering en manipulatie van deze sociale grafen heeft de belangstelling gewekt vanuit verschillende onderzoeksgemeenschappen. Hierdoor is veel vooruitgang geboekt in het analyseren van sociale grafen. Er is echter een nog snellere vooruitgang geboekt in data verzamelingstechnieken, wat geresulteerd heeft in een grotere hoeveelheid data, meer verschillende soorten data en minder waarheidsgetrouwheid in de beschikbare bronnen.

Met de opkomst van sociale netwerken en de breed toegankelijke historische archieven, hebben de complexiteit en dynamica van de resulterende sociale grafen geleid tot de volgende twee onderzoeksvragen. Ten eerste, hoe kan de structuur en dynamiek uit de sociale heterogene data gehaald worden en worden omgezet in grafen, oftewel hoe kan sociale data worden weergegeven als een sociale graaf. Ten tweede, hoe kunnen de eigenschappen van dynamische sociale grafen, ofwel evoluerende sociale grafen, gemodelleerd en geanalyseerd worden.

De analyse van dynamische sociale grafen is een overkoepelend thema in dit proefschrift. Hoofdstuk 2 geeft de benodigde achtergrondkennis betreffende *information retrieval*, sociale grafen en de theorie van dynamische systemen, die nodig is om dit proefschrift beter te kunnen begrijpen. Verder worden hier de datasets gintroduceerd die op verschillende plaatsen in dit proefschrift worden gebruikt voor experimentele analyses en numerieke verificatie. De data-gedreven benaderingen, voorgesteld in Hoofdstuk 3 en 4, leiden tot krachtige praktische gereedschappen voor de integratie van heterogene sociale databestanden. Deze gereedschappen resulteerden in het automatisch genereren van een set van verrijkte sociale grafen, welke met bestaande analytische methoden niet geanalyseerd konden worden.

Twee nieuwe analytische modellen worden voorgesteld in Hoofdstuk 5 die een diepgaand begrip van de topologische eigenschappen van sociale grafen mogelijk maken. In hoofdstuk 6 wordt een nieuw analytisch model gintroduceerd waarmee de evolutie

van gedrag in zulke grafen kan worden gemodelleerd. Dit model beschrijft met succes verschillende belangrijke eigenschappen van dynamische sociale grafen. Een volgende onderzoeksstap wordt genomen in Hoofdstuk 7: de gelijktijdige evolutie van topologie en gedrag in sociale grafen wordt bestudeerd, en het effect van deze co-evolutie op de topologische diversiteit en de evenwichtstoestanden van de sociale grafen wordt aangetoond.

Concluderend worden er in dit proefschrift effectieve methoden ontwikkeld om heterogene sociale data te verwerken tot sociale grafen en worden er nieuwe modellen voorgesteld ter analyse van deze dynamische sociale grafen. Uit dit proefschrift volgen ten minste vier nieuwe, veelbelovende onderzoeksrichtingen. Ten eerste maakt dit *Geavanceerd Genealogisch Onderzoek* mogelijk. Ten tweede paveit dit de weg voor het *Traceren van de Oorsprong van 'Heavy-Tailed' Verdelingen* in sociale netwerken door middel van dynamische-systeemtheorie. Ten derde levert het nieuwe methoden op voor het bestuderen van de *Regelbaarheid van Gedragingen in Sociale Netwerken*. Ten vierde wordt er een brug geslagen tussen de vakgebieden van de *Zwermrobotica* en de *Experimentele Economie*.