Heterogeneous Graphs in Social Business
Kush R. Varshney, Jun Wang, and Aleksandra Mojsilović
Agenda

Brief introduction to business analytics

Social business and connections

Constrained graph drawing for visualization

Link prediction on a heterogeneous graph

Conclusion
Some businesses operate in the same way that they always have

http://www.telegraph.co.uk/finance/newsbysector/retailandconsumer/8114961/Alliance-Boots-plans-major-push-into-China.html
Some businesses operate in the same way that they always have.
Some businesses operate in the same way that they always have
Some businesses now collect and examine some data
Some businesses are integrating data-driven algorithmic recommendations into their decision-making processes

- These organizations are adopting *business analytics*

- Business analytics is all about the decisions that go into running a business
  - Not quantitative finance or investing in stocks
  - Things like:
    - Hiring, retention, attrition
    - Buying, selling, marketing
    - Staff deployment
    - Strategy
What is business analytics

- Encompasses methodologies from applied mathematics, applied probability, applied statistics, computer science, and signal processing for using data to gain insight into business performance and drive business planning.

- Solutions primarily used as decision support systems or as components of decision support systems to aid salespeople, executives, and other organizational leaders in business decision-making tasks.

Business decisions

- How likely is client X to buy product Y?
- Which product should we recommend next?
- What is a “realistic” view of opportunity by client?
- Are there accounts where there is significant untapped revenue?
- Which clients are “at risk” of going to our competitors?
- What kind of salesforce do we need to deliver on our targets?
- Which units are not performing at par?
- Which sellers might miss their quota?
- How can we align sellers with client opportunities to achieve maximum revenue impact?
Business decisions (continued)

- Who are the influencers for this product in the marketplace?
- What is the optimal marketing campaign to deploy?
- Should we hire this employee?
- Which employees are at risk of voluntarily leaving the company?
- How many employees do we need to hire now so that we can achieve our production goals six months from now?
- What kind of raises or promotions should we offer to retain our talented employees?
- Will outsourcing help?
- Which company should we acquire to expand our customer base?
Agenda

Brief introduction to business analytics

Social business and connections

Constrained graph drawing for visualization

Link prediction on a heterogeneous graph

Conclusion
Social business

- Much of business revolves around who you know and what you know
- 2012 is being called ‘the year of social business’
- Connections and networking
- All sorts of links
  - Companies connected to other companies
  - People connected to other people
  - People connected to companies
  - Companies connected to products
  - People connected to products
  - People connected to knowledge
  - Knowledge connected to products
Market social network

- Network that captures the relationships among different companies in the marketplace
- The attributes for each node include information such as firmographics or keywords extracted from company web pages or news pages
- The edges among different companies include categories such as business transactions (who bought what from whom), business alliances (joint ventures, partnerships), officer relations, similar culture (‘green’ companies, ‘growing’ companies)
- Structured and unstructured data sources
  - Firmographics: Dun & Bradstreet, etc.
  - Financial performance: S&P Compustat (for publicly traded companies)
  - Buying relationships: corporate customer relationship management systems, company press releases, corporate market intelligence
  - Executive relationships: Boardex, news (Factiva, Reuters, etc.)
  - Social media analysis
Examples of board and executive links

- First-degree connection

<table>
<thead>
<tr>
<th>Organization</th>
<th>Start Date</th>
<th>Role</th>
<th>Person</th>
<th>Start Date</th>
<th>Role</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>CREDIT AGRICOLE SA</td>
<td>22 Jun 2009</td>
<td>Advisor (Non-Brd)</td>
<td>The Hon. Doctor Kwok-Po (David)</td>
<td>Li</td>
<td>External Director (Brd - SD)</td>
<td>CAIX ABANK SA</td>
</tr>
</tbody>
</table>

- Second-degree connection

<table>
<thead>
<tr>
<th>Organization</th>
<th>Start Date</th>
<th>Role</th>
<th>Person</th>
<th>Start Date</th>
<th>Role</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMMERZBANK AG</td>
<td>01 Apr 2004</td>
<td>CFO (Brd - ED)</td>
<td>Doctor Eric Wolfgang Strutz</td>
<td></td>
<td>Director - SD (Brd - SD)</td>
<td>MEDICANSA SPA</td>
</tr>
<tr>
<td></td>
<td>Current</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26 Oct 2008</td>
<td>Director - SD (Brd - SD)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Antoine M Berrisheim</td>
<td></td>
<td>Board Member - SD (Brd - SD)</td>
<td>GENERALI DEUTSCHLAND MÖLDING AŞ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1998</td>
<td>Current</td>
<td></td>
</tr>
</tbody>
</table>

© 2012 IBM Corporation
The social network can be visually displayed to help marketing and sales understand the risks and opportunities among a certain set of accounts.

Highlight companies that are not currently customers (whitespace) that could become customers.
Identification of sales experts within a company

- Enterprise + market social network: which sellers should staff specific accounts to yield higher revenues
- Can also have products as a third type of node
Company acquisition

- Which acquisition target will expand the company’s customer base into whitespace

(customer node size indicates revenue)
Company acquisition (continued)

- Which acquisition is better?

A has good whitespace coverage. It also sells to many of our clients. Its whitespace accounts have similar characteristics to existing IBM accounts. This indicates that A’s whitespace accounts could have high propensity to buy other products/services from IBM.

D has almost no whitespace coverage, primarily serving several of our key accounts.

B has good whitespace coverage. It also sells to many of competitor’s clients. B could be the acquisition of choice if we are considering penetrating into competitive accounts.

C has good whitespace coverage. Its clients however, do not seem to be buying any other products (IBM or competitive) indicating low propensity to grow revenue in other businesses.

Customer node size indicates revenue.
What pieces of marketing collateral should be recommended to sellers?

Agenda

Brief introduction to business analytics
Social business and connections
Constrained graph drawing for visualization
Link prediction on a heterogeneous graph
Conclusion
Graph drawing problem

- Given a set of nodes connected by weighted edges, determine the placement of the nodes in the 2-dimensional plane
  - Distance in drawing should be related to edge weight
  - Different specific objectives and algorithms

- Spectral graph drawing (Hall, 1970)
  - use eigenvectors of Laplacian of adjacency matrix

- Degree-normalized eigenvectors are more aesthetically pleasing
Spectral graph drawing

- Adjacency matrix $A$
- Degree matrix $D$ is a diagonal matrix containing row sums of $A$
- Laplacian matrix $L = D - A$
- Find generalized eigenvectors of $(L, D)$
- Coordinates of nodes are elements of the eigenvectors corresponding to the $2^{nd}$ and $3^{rd}$ smallest eigenvalues
- Eigenprojection motivated by energy-minimization objective
- Does not allow specific nodes to be placed in a certain way
  - Treat vendor nodes differently
    - Constrain to sides of visualization
    - Free placement on unit circle
    - Fixed placement, e.g. on left and right for $n = 2$

Power iteration with constrained nodes

- Computing the eigendecomposition
- Power iteration (Golub and van Loan, 1996)

- Constraining nodes
- Free placement on unit circle
  - extra normalization $\|[\hat{u}^2_j \hat{u}^3_j]^T\| = 1$ for all $j$ vendor nodes
- Fixed placement
  - reset values $[\hat{u}^2_j \hat{u}^3_j]$ on every iteration for the nodes whose placement is to be fixed

Marketplace visualization

Marketplace visualization (zoomed in)

Agenda

Brief introduction to business analytics
Social business and connections
Constrained graph drawing for visualization
Link prediction on a heterogeneous graph
Conclusion
Link prediction on a heterogeneous graph

- Novel random walk model-based inference with complex and dynamic graph structures
- Complex
  - Heterogeneous graph containing multiple types of nodes and edges, including social connections
- Dynamic
  - Random walk-based algorithm adds new connections dynamically
- Predict edges to recommend new relevant connections

Illustration of predictive formulation

Generating predictions for decision support

(a) **Perform random walk over unipartite graphs to derive stationary distributions (static)**

\[ R(t + 1) = \alpha PR(t) + (1 - \alpha)I \]

(b) **Iteratively perform two-way random walk across the unipartite graphs (dynamic)**

\[ P_{(xy)}^{yea}(t + 1) = \gamma R(x)P_{(xy)}^{yea}(t) + (1 - \gamma)P_{(xy)}^{yea}(t)R(y) \]

Illustration of predictive formulation

Generating predictions for decision support
(a) Perform random walk over unipartite graphs to derive stationary distributions (static)
\[ R(t + 1) = \alpha PR(t) + (1 - \alpha)I \]
(b) Iteratively perform two-way random walk across the unipartite graphs (dynamic)
\[ P^{yea}_{(xy)}(t + 1) = \gamma R(x)P^{yea}_{(xy)}(t) + (1 - \gamma)P^{yea}_{(xy)}(t)R(y) \]

Generating predictions for decision support

(a) Perform random walk over unipartite graphs to derive stationary distributions (static)
\[ R(t + 1) = \alpha PR(t) + (1 - \alpha)I \]

(b) Iteratively perform two-way random walk across the unipartite graphs (dynamic)
\[ P^{\text{yea}}_{(xy)}(t + 1) = \gamma R(x)P^{\text{yea}}_{(xy)}(t) + (1 - \gamma)P^{\text{yea}}_{(xy)}(t)R(y) \]

Generating predictions for decision support

(a) Perform random walk over unipartite graphs to derive stationary distributions (static)

\[ P_{(x\rightarrow y)}(t + 1) = \gamma R(x) P_{(x\rightarrow y)}(t) + (1 - \gamma) R(y) \]

(b) Iteratively perform two-way random walk across the unipartite graphs (dynamic)

\[ R(t + 1) = \alpha P R(t) + (1 - \alpha) I \]

Illustration of predictive formulation

Generating predictions for decision support

(a) Perform random walk over unipartite graphs to derive stationary distributions (static)

$$R(t + 1) = \alpha PR(t) + (1 - \alpha)I$$

(b) Iteratively perform two-way random walk across the unipartite graphs (dynamic)

$$P^{yea}_{(xy)}(t + 1) = \gamma R(x) P^{yea}_{(xy)}(t) + (1 - \gamma) P^{yea}_{(xy)}(t) R(y)$$

Illustration of predictive formulation

Generating predictions for decision support

(a) *Perform random walk over unipartite graphs to derive stationary distributions* (**static**)

\[ R(t + 1) = \alpha PR(t) + (1 - \alpha)I \]

(b) *Iteratively perform two-way random walk across the unipartite graphs* (**dynamic**)

\[ P_{(xy)}^{yes}(t + 1) = \gamma R_x P_{(xy)}^{yes}(t) + (1 - \gamma)P_{(xy)}^{yes}(t)R(y) \]

Illustration of predictive formulation

Generating predictions for decision support

(a) Perform random walk over unipartite graphs to derive stationary distributions (static)

\[ R(t + 1) = \alpha P R(t) + (1 - \alpha) I \]

(b) Iteratively perform two-way random walk across the unipartite graphs (dynamic)

\[ P_{(xy)}^{yea}(t + 1) = \gamma R(x) P_{(xy)}^{yea}(t) + (1 - \gamma) P_{(xy)}^{yea}(t) R(y) \]

(c) Predict new edges based on information-theoretic criterion

Agenda

Brief introduction to business analytics
Social business and connections
Constrained graph drawing for visualization
Link prediction on a heterogeneous graph

Conclusion
Conclusion

- Businesses are increasingly turning to data-driven algorithmic insights and recommendations to support decisions
- Much of business is social and connections-based
  - Various, wide-ranging problems and tasks
- Highlighted two graph-based mathematical formulations
  - Constrained graph drawing
  - Random walk-based edge prediction in a heterogeneous graph
Questions