FINDING MEANING IN SOCIAL MEDIA: CONTENT-BASED SOCIAL NETWORK ANALYSIS OF QUITNET TO IDENTIFY NEW OPPORTUNITIES FOR HEALTH PROMOTION

Sahiti Myneni\textsuperscript{a,b}, Nathan Cobb\textsuperscript{c}, Trevor Cohen\textsuperscript{a}

\textsuperscript{a} National Center for Cognitive Informatics and Decision Making in Healthcare, UT School of Biomedical Informatics at Houston
\textsuperscript{b} Innovations in cancer Prevention Research, School of Public Health, The University of Texas Health Science Center at Houston, TX, USA
\textsuperscript{c} The Schroeder Institute for Tobacco Research and Policy Studies, Washington, DC, USA
AGENDA

Introduction & Background
• Health behaviors
• Role of social networks
• Current state of online social network analysis

Research Methodology
• Qualitative analysis
• Automated text analysis
• Network analysis

Results
Innovation
Limitations
Conclusions & Implications
HEALTH BEHAVIORS

Preventable illness (63% of global deaths)

• Tobacco use
• Poor diet
• Physical inactivity
• Harmful alcohol use

Modifiable health behaviors account for 30% of cancer related deaths
CURRENT STATE OF INTERVENTIONS

STICK APPROACH
CARROT APPROACH
Social relationships and health behaviors
(Valente, 2010; Christakis & Fowler, 2008)

• Homophily (McPherson et al., 2001)
• Observational learning (Bandura, 1986)
• Social support (House, 1981)

Online Social Networks

• Data source AND intervention delivery platforms
• Real-time, real-life (Siffman et al., 2010)
CONTENT & NETWORKS

Distributional Semantics

• Hyperspace Analogue to Language (HAL) in online social networks (McArthur et al., 2006)
• Structure-free

Semantic Relatedness

• Pathfinder network (Schvaneveldt, 1990)
• Content determines structure
### Current paradigm
Communication frequency

<table>
<thead>
<tr>
<th>Sender ID</th>
<th>Receiver ID</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>22</td>
<td>Great 1 week of abstinence, keep going, congratulations</td>
</tr>
<tr>
<td>22</td>
<td>11</td>
<td>I had a rough day, lots of tension and craving for nicodemon</td>
</tr>
</tbody>
</table>

### New paradigm
Communication frequency and content

<table>
<thead>
<tr>
<th>Sender ID</th>
<th>Receiver ID</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>88</td>
<td>55</td>
<td>I understand what you are going through, please don’t give up</td>
</tr>
<tr>
<td>55</td>
<td>22</td>
<td>I extend my hand to you for a smokeless day</td>
</tr>
<tr>
<td>66</td>
<td>11</td>
<td>Gaining weight is normal when you quit smoking</td>
</tr>
</tbody>
</table>
QuitNet, an online social network for smoking cessation

- 100,000 new registrants/year
- Participation strongly correlated with abstinence (Cobb et al, 2008)

A database of 16,492 de-identified messages between March 1, 2007 and April 30, 2007
RESEARCH STRATEGY

Thematic analysis to identify behavioral components

Distributional semantics and vector space modeling

Social network modeling

Content-specific intervention personalization
QUALITATIVE ANALYSIS

100 randomly chosen messages

• Behavioral constructs, context

• Alignment with behavior change theories
  • Social Cognitive Theory (Bandura, 1986)
  • The Transtheoretical Model of Change (Prochaska & Velicer, 1997)

• Grounded Theory approach (Strauss & Corbin, 1990)
  • Open coding
  • Axial coding
  • Constant comparison
Thanks for hosting the bonfire tonight. I have been dragging around this lawnbag of 750 unsmoked cigarettes, they have just been piling up over the past 25 days. I would like to donate them to the bonfire.

It helped so much to hear from all of you. I have tears of relief right now that I feel in control again.
AUTOMATED ANALYSIS

- Latent Semantic Analysis (LSA) using Semantic Vectors package (Widdows & Ferraro, 2008)

- LSA derives relatedness measures between terms or passages from unannotated text by representing the terms in a high-dimensional vector space

1. TASA corpus
2. QuitNet corpus
3. Term vectors
4. Message vectors
5. Pair vectors
6. Theme Vectors
7. Semantic similarity score

From Qualitative analysis:
- sad emotion help

Distributional statistics
RESULTS: AUTOMATED ANALYSIS

Theme-specific similarity score for each pair of QuitNet users

Validation: Five random QuitNet users chosen

- 82 messages, 27 unique users
- The messages exchanged were manually coded
- Recall~0.75
  - messages retrieved/messages discussing theme
- Precision~0.81
  - relevant messages retrieved/total messages retrieved
CONTENT-BASED NETWORK MODELING

- Users $\rightarrow$ nodes; Communication between users $\rightarrow$ edges
- Statistical threshold to limit the links (Mean+Std Dev)
- Gephi, an open source network analysis and visualization software

<table>
<thead>
<tr>
<th>User pairs exchanging messages</th>
<th>Reinforcement TH=0.787</th>
<th>Personal Experience TH=0.624</th>
</tr>
</thead>
<tbody>
<tr>
<td>1616971 1056108</td>
<td>0.778</td>
<td>0.517</td>
</tr>
<tr>
<td>1142992 1616971</td>
<td>0.541</td>
<td>0.399</td>
</tr>
<tr>
<td>1567343 1631267</td>
<td>0.819</td>
<td>0.624</td>
</tr>
<tr>
<td>1123493 1056108</td>
<td>0.657</td>
<td>0.644</td>
</tr>
<tr>
<td>1302604 1418369</td>
<td>0.858</td>
<td>0.588</td>
</tr>
<tr>
<td>1750563 1080691</td>
<td>0.788</td>
<td>0.536</td>
</tr>
<tr>
<td>1248151 1615961</td>
<td>0.614</td>
<td>0.479</td>
</tr>
</tbody>
</table>
RESULTS: CONTENT-SPECIFIC NETWORK MODELS

- Network density and structure varied across themes
- High-degree nodes were different
- Clustering of opinion leaders within each network
Reinforcement  
Support  
Personal Experience
LIMITATIONS

• QuitNet dataset recorded in 2007, limited in size
• Qualitative analysis until thematic saturation
• Evaluation framework can be more robust including multiple users
INNOVATION

• Facilitates content inclusion in network science
• Effectively merges qualitative, automated, and quantitative methods
• Extends the scalability of qualitative methods
• Demonstrates the translational design of behavior change support interventions
• Provides granular and comprehensive view of human behavior
CONCLUSIONS & NEXT STEPS

• Modifiable health behaviors contribute to the majority of deaths globally

• Online social networks provides global perspective and platform to investigate and intervene

• A “structure plus content” method to understand inter- and intra-individual behavioral intricacies

• Personalized and targeted solutions to initiate or adhere to a positive behavior change

• Translational interventions that harness the power of social relationships
ACKNOWLEDGEMENT

• Cancer Prevention Research Institute of Texas, USA (CPRIT)

• Executive committee, mentors, and fellows of CPRIT Innovations Training Grant, School of Public Health, The University of Texas Health Science Center, Houston, TX, USA

• School of Biomedical Informatics

• QuitNet
THANK YOU
REFERENCES