Determining Parkinson’s Sub-types through Trajectory Clustering in Bipartite Networks

Sanjukta Krishnagopal, Michelle Girvan
University of Maryland, College Park

Collaborators: Rainer von Coelln, Lisa Shulman,
University of Maryland School of Medicine, Baltimore

NetSci 2018
Motivation

- Parkinson’s disease (PD) is a heterogenous progressive neurodegenerative condition. 7 to 10 million people worldwide are living with Parkinson's.

- PD symptoms/variables—motor, cognitive, psychiatric etc.—vary significantly among individuals, suggesting disease subtypes.

- There’s a lack of reliable, objectives measures of PD sub-type and accurate early-assessment of it.

- Only recently do we have systematically collected large-scale data and the necessary computational power to use data-driven methods in network science for PD.
Parkinson’s Progressive Markers Initiative (PPMI) sponsored by The Michael J. Fox Foundation.

Data over 4 years of 17 PD clinical variables (cognitive, behavioural etc.) + genetic alleles of ~400 recently diagnosed PD and ~200 healthy subjects.

- Identify groups of variables/indicators that co-occur through community detection.
- Identify ‘typical’ trajectories that patients take through disease progression.
- Develop a Trajectory Clustering algorithm to identify PD patient sub-types through trajectory clustering.
- With knowledge of the patient’s initially disease profile, we can predict disease trajectory and PD subtype assisting in early detection personalized medicine.

www.ppmi-info.org/data
# PPMI Data - PD Variables

<table>
<thead>
<tr>
<th>Domain</th>
<th>Test</th>
<th>Code</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PD Rating</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD1</td>
<td>Unified Parkinson’s Disease Rating Scale - Part 1</td>
<td>UPDRS 1</td>
<td>Mentation, Behavior and Mood</td>
</tr>
<tr>
<td>PD2</td>
<td>Unified Parkinson’s Disease Rating Scale - Part 2</td>
<td>UPDRS 2</td>
<td>Activities of Daily Life</td>
</tr>
<tr>
<td>PD3</td>
<td>Unified Parkinson’s Disease Rating Scale - Part 3</td>
<td>UPDRS 3</td>
<td>Clinician-Scored Monitored Motor Evaluation</td>
</tr>
<tr>
<td>PDT</td>
<td>Total Unified Parkinson’s Disease Rating Scale</td>
<td>T-UPDRS</td>
<td>Total</td>
</tr>
<tr>
<td><strong>Autonomic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td>SCales for Outcomes in PArkinson’s disease</td>
<td>SCOPA-AUT</td>
<td>Autonomic Dysfunction</td>
</tr>
<tr>
<td><strong>Cognitive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>Hopkins Verbal Learning Test</td>
<td>HVLT</td>
<td>Memory</td>
</tr>
<tr>
<td>C2</td>
<td>Benton Judgment of Line Orientation</td>
<td>JOLO</td>
<td>Visuospatial Skills</td>
</tr>
<tr>
<td>C3</td>
<td>Senior Fitness Test</td>
<td>SFT</td>
<td>Functional Fitness</td>
</tr>
<tr>
<td>C4</td>
<td>Letter-Number Sequencing</td>
<td>LNS</td>
<td>Working memory</td>
</tr>
<tr>
<td>C5</td>
<td>Symbol-Digit Modalities Test</td>
<td>SDM</td>
<td>Attention</td>
</tr>
<tr>
<td>C6</td>
<td>Montreal Cognitive Assessment</td>
<td>MOCA</td>
<td>Memory, Visuospatial, Attention, Concentration, Language</td>
</tr>
<tr>
<td><strong>Sleep</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>Epworth Sleepiness Scale</td>
<td>ESS</td>
<td>Daytime sleepiness</td>
</tr>
<tr>
<td>S2</td>
<td>REM Sleep Behavior Disorder Questionnaire</td>
<td>RBDQ</td>
<td>RBD</td>
</tr>
<tr>
<td><strong>Behavior</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>Geriatric Depression Scale</td>
<td>GDS</td>
<td>Depression</td>
</tr>
<tr>
<td>B2</td>
<td>State-Trait Anxiety</td>
<td>STAI</td>
<td>Anxiety</td>
</tr>
<tr>
<td><strong>Disability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dis</td>
<td>Schwab &amp; England Activities of Daily Living Scale</td>
<td>SEADL</td>
<td>Ability to Function in Activities of Daily Living</td>
</tr>
<tr>
<td><strong>Gene</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>rs11060180</td>
<td>G1</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>rs6430538</td>
<td>G2</td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>rs823118</td>
<td>G3</td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>rs356181</td>
<td>G4</td>
<td></td>
</tr>
<tr>
<td><strong>Gender, Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bipartite Representation and Community Detection

- **Bipartite network** of patients and variables/indicators for each layer (each time-point or specific ranges of disease severity):

- Patient \( i \) is connected to variable \( v \) with a \( z \) score: 
  \[
  z_{iv} = \frac{X_{iv} - \langle X^b \rangle_v}{\sigma^b_v}
  \]
  where superscript \( b \) denotes baseline year.

- **Community Detection** on the bipartite network\(^1\) for each layer.

---

• **Bipartite network** of patients and variables/indicators for each layer (each time-point or specific ranges of disease severity):

• Patient $i$ is connected to variable $v$ with a $z$ score:  
  \[ z_{iv} = \frac{X_{iv} - <X^{b}>_v}{\sigma^{b}_v} \]
  where superscript $b$ denotes baseline year.

• **Community Detection** on the bipartite network\(^1\) for each layer.

• Stack communities in each layer to create a **stacked multi-layer network**.
  • Track patient trajectories through variable communities (nodes) across time.

Bipartite Representation and Community Detection

- **Bipartite network** of patients and variables/indicators for each layer (each time-point or specific ranges of disease severity):

- Patient $i$ is connected to variable $v$ with a $z$ score: $z_{iv} = \frac{X_{iv} - \langle X^b \rangle_v}{\sigma^b_v}$ where superscript $b$ denotes baseline year.

- **Community Detection** on the bipartite network\(^1\) for each layer.

- Stack communities in each layer to create a **stacked multi-layer network**.
  - Track patient trajectories through variable communities (nodes) across time.

---

Develop our **second-order Trajectory Clustering Algorithm**:

- Define **node closeness** $cl(x,y)$ between two nodes $x$ and $y$ as:

  $$cl(x, y) = \sum_t \frac{A_t \cap B_t}{A_t \cup B_t}$$

  $A_t, B_t$ are edges from layer $t$ reached through trajectories from nodes $x, y$ respectively.

- Patient-Patient Adjacency matrix given by **Trajectory Similarity (TS)**. For two patients $i, j$ with trajectories $T_i$ and $T_j$:

  $$A_{i,j} = A_{T_i,T_j} = \sum_{x,y} cl(x_{T_i}, y_{T_j})$$

- **Louvain community detection**\(^2\) on the patient-patient network.

Trajectory Clustering with Temporal Layers

Number of people in each community:
- **red**: 41
- **blue**: 64
- **green**: 56
- **yellow**: 69

Four distinct trajectory clusters:
- **red**: sleep
- **blue**: behavior and UPDRS
- **green**: age, cognitive 1
- **yellow**: male, cognitive 2
Trajectory Clustering with Disease Progression Layers

Layers represents disease progression measured via variable UPDRS3.

Three distinct trajectory clusters:

**green:** start healthy, limited progression (59 people)

**red:** start with high disease activity, exhibit Cognitive 1 variables (20 people)

**blue:** high disease activity, substantial progression, exhibit Cognitive 2 variables (56 people)
Trajectory Clustering in Variable Profile Space

- What if variables don’t cluster naturally (low modularity)?
  Track patients through a more general unclustered variable space.

- A patient $i$’s trajectory profile is represented by a matrix $(T_i)$ of size $V \times Y$ where $V$ is the total number of variables and $Y$ is number of time-points.

- This emphasizes a different aspect of Trajectory Similarity where variables are treated as independent as opposed to being clustering into communities.

- We include genetic data in order to study genotype-phenotype relationships.
Patient communities based on Variable Profile Trajectory Clustering

- Number of people
- Time

Domains:
- Gender-Male
- Age - Older
- Cognitive
- Disability
- Sleep
- Autonomic
- Behavior
- UPDRS

Groups:
Group 1
Group 2
Group 3
Patient communities based on Variable Profile Trajectory Clustering

Number of people

95

43

102

Domains

Gender-Male

Age - Older

Cognitive

Disability

Sleep

Autonomic

Behavior

UPDRS

Gene

Normalized variable intensity

High

Low

Group 1:
- Initially low then progressive UPDRS
- high Cognitive 1
- high Gene 3_CC variant

Group 2:

Group 3:
- Youngest cluster.
- Initially low then slowly progressing UPDRS, high Cognitive 2, high Gene 1_CT, Gene 2_TT, Gene 3_TT, Gene 4_CC variants, low Gene 3_CC variant.
Patient communities based on Variable Profile Trajectory Clustering

Number of people
95
43
102

Domains
- Gender: Male, Older
- Cognitive
- Disability
- Sleep
- Autonomic
- Behavior
- UPDRS
- Gene

Group 1:
- Initially low then progressive UPDRS
- high Cognitive 1
- high Gene 3_CC variant

Group 2:
- Initially moderate then progressive UPDRS
- high Cognitive 1
- unremarkable Gene profile

Group 3:
- Youngest cluster.
- Initially low then slowly progressing UPDRS, high Cognitive 2, high Gene 1_CT, Gene 2_TT, Gene 3_TT, Gene 4_CC variants, low Gene 3_CC variant.
Patient communities based on Variable Profile Trajectory Clustering

- **Group 1:**
  - Initially low then progressive UPDRS
  - high Cognitive 1
  - high Gene 3_CC variant

- **Group 2:**
  - Initially moderate then progressive UPDRS
  - high Cognitive 1
  - unremarkable Gene profile

- **Group 3:**
  - Youngest cluster.
  - Initially low then slowly progressing UPDRS
  - high Cognitive 2
  - high Gene 1_CT, Gene 2_TT, Gene 3_TT, Gene 4_CC variants
  - low Gene 3_CC variant
Conclusion

- Patient disease-trajectory prediction and sub-typing that can be generalized to any progressive multi-variate disease.

- Identify groups of variables/indicators that co-occur in different stages of disease progression.

- Develop a Trajectory Clustering algorithm to cluster patient trajectories in order to find typical patient sub-types.

- **Strengths:** Data-driven method - useful in early-prediction personalized medicine.

- **Limitations:** Must be used in conjunction with clinical expertise.

- This network analytics is part of a project that aims to integrate the analytics into a disease visualization tool directed at the medical community called BioViz™.
Questions?