Network science is interdisciplinary. Focusing on improving current analytical strategies, I try to absorb influences of other disciplines to inform and improve understandings about human behaviors with applications in neuroscience, psychiatry, psychology and others. Broadly speaking, I model real-world networks with statistically sound principles and methods—reduce network complexity, perform intuitive visualization and test critical hypotheses with improved statistical power and controlled type I errors. I am particularly interested in the topological structures of networks and the roles they play in (neuro)degeneration, (neuro)development and resilience.

1 Brain Networks

LatentSNA. To improve the statistical power for understanding the brain-to-behavior relationship, we propose a latent variable-based statistical network analysis (LatentSNA) that combines brain **functional connectivity** with internalizing psychopathology, implementing network science in a generative statistical process to preserve the neurologically meaningful network topology in the adolescents and children population. The developed inference-focused generative Bayesian framework (1) addresses the lack of power and inflated Type II errors in current analytic approaches when detecting imaging biomarkers, (2) allows unbiased estimation of biomarkers' influence on behavior variants, (3) quantifies the uncertainty and evaluates the likelihood of the estimated biomarker effects against chance and (4) ultimately improves brain-behavior prediction in novel samples and the clinical utilities of neuroimaging findings. We collectively model multi-state functional networks with multivariate internalizing profiles for **5,000 to 7,000 children in the Adolescent Brain Cognitive Development (ABCD) study** with sufficiently accurate prediction of both **children internalizing** traits and functional connectivity, and substantially improved ability to explain the individual internalizing differences over current approaches. This work is submitted to *Nature Communications*.

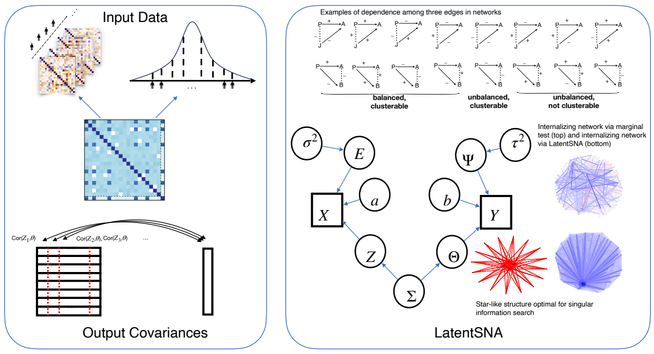

Figure 1: The schematic diagram of LatentSNA.

ABC. Brain structural connectivity, capturing the **white matter fiber tracts among brain regions inferred by diffusion MRI (dMRI)**, provides a unique characterization of brain anatomical organization. One fundamental question to address with structural connectivity is how to properly summarize and perform statistical inference for a group-level connectivity architecture, for instance, under different sex groups, or disease cohorts. Existing analyses commonly summarize group-level brain connectivity by a simple entry-wise sample mean or median across individual brain connectivity matrices. However, such a heuristic approach fully ignores the associations among structural connections and the topological properties of brain networks. In this project, we propose a latent space-based generative network model to estimate group-level brain connectivity. Within our modeling framework, we incorporate the anatomical information of brain regions as the attributes of nodes to enhance the plausibility of our estimation and improve biological interpretation. We name our method the attributes-informed brain connectivity (ABC) model, which compared with existing group-level connectivity estimations, (1) offers an interpretable latent space representation of the group-level connectivity, (2) incorporates the anatomical knowledge of nodes and tests its co-varying relationship with connectivity and (3) quantifies the uncertainty and evaluates the likelihood of the estimated group-level effects against chance. We devise a novel Bayesian MCMC algorithm to estimate the model. We evaluate the performance of our model through extensive simulations. By applying the ABC model to study brain structural connectivity stratified by sex among **Alzheimer's Disease** (AD) subjects and healthy controls incorporating the anatomical attributes (volume, thickness and area) on nodes, our method shows superior predictive power on out-of-sample structural connectivity and identifies meaningful sex-specific network neuromarkers for AD. This work is invited for revision at *Medical Image Analysis*.

2 Social Networks

JLSM. A joint modeling approach implies that both network data and item response data are modeled as stemming from a shared data generation process, which comes with a few benefits. I propose a joint latent space model (JLSM) using a shared latent variable to explain the dependence between two sources of data.[1] **JLSM devises an interactive joint latent space, where two types of nodes (person and item nodes) can coexist.** By visualizing the joint latent space of French Elites, I discovered a division in their social circles following patterns of political, educational, and social class differences. In an application to the classroom educational outcomes and students' social network,[2] I identified isolated and under-achieving students through a joint visualization of students' social relationships and reports of academic outcomes. **JLSM poses as a powerful visualization and prediction tool while accounting for dependencies within and between social networks and item responses (see applications in Figure 2), which was not possible until now.** This work is accepted at *Psychometrika*.

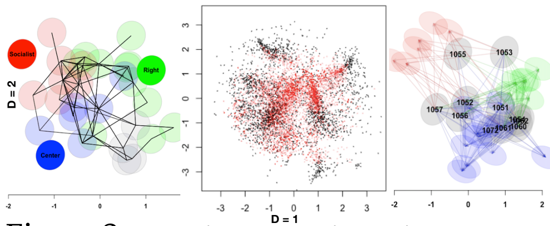

Figure 2: Two-dimensional joint latent spaces for French elites with opposing party affiliations (left), online YouTube users (middle) and students' attitudes towards school (right).

Software. I developed a novel **Variational Bayesian Expectation-Maximization algorithm for estimating Euclidean latent space models for bipartite networks**. The algorithm accommodates **large-scale datasets** and is documented in the jlsm package that I made,[3] which is freely available on CRAN and has several modeling approaches.

JNIRM. I proposed an alternative joint modeling framework using a correlational link between the network and the latent variable item response model (JNIRM). The JNIRM framework makes two key contributions: first, it presents a novel strategy allowing **simultaneous inference between and within social networks and item responses**. Second, it allows **testing of dependencies between dimensions of latent (psychological) constructs and dimensions of latent network space**. Neither has previously been accomplished by existing research. In an application to the teachers' advising relationships and their job perceptions,[4] we found that flourishing advising relationships contribute to a more satisfying teaching environment; in addition, we observe that the complementarity principle applies, where teachers tend to seek advice from those who are different from them. This work is under revision.

3 Bipartite Networks and Multivariate Analysis

Bipartite Networks. In a paper accepted at *Journal of Complex Networks*, we applied the latent position cluster model to a bipartite Islamic State recruitment network to investigate the **terrorist organization's resilience in the face of external stress.**[5] We compared the recruitment network to a hypothetical organization, where recruitments were independently generated without assuming dependence or association between nodes, and we found that **the Islamic organization was much more resilient than the forementioned hypothetical organization, which partly explained why decapitation strikes against terrorist organizations were not as effective as expected**.

Psychometrics. I focus on understanding the significance of reliability. In a paper accepted at *Behavior Research Methods*, I aired caution about interpreting reliability coefficients in individual experimental studies.[6] In a between-subjects study, **reliability coefficients do not simply indicate a quality aspect of measures, but also denote the heterogeneity of the sample**. This understanding of reliability has complications for its relationship with statistical power and can lead to seemingly **paradoxical observations that a low reliability can co-occur with high statistical power**. The theoretical results explain observations from **large-scale replication studies**, where negative correlations are found between reliability coefficients and power.

Latent Factor Models. In a paper accepted at *Applied Psychological Measurement*, we investigated improper solutions (negative variance estimates, Heywood cases), convergence issues, and outlying estimates in **factor analytic models**.[7] The problem can be cast in **item response theory** frameworks as parameters of the two models are related given certain constraints (see Equations above). Heywood cases with delta parametrizations show as nonconvergent issues if theta parameterizations are used and as outlying discriminations with an IRT approach. The results provide insights for researchers who encounter estimation problems. A follow-up manuscript is in progress about the causes of Heywood cases, which provides associated remedies for underlying patterns of data that signal potential estimation problems.

4 Future Directions

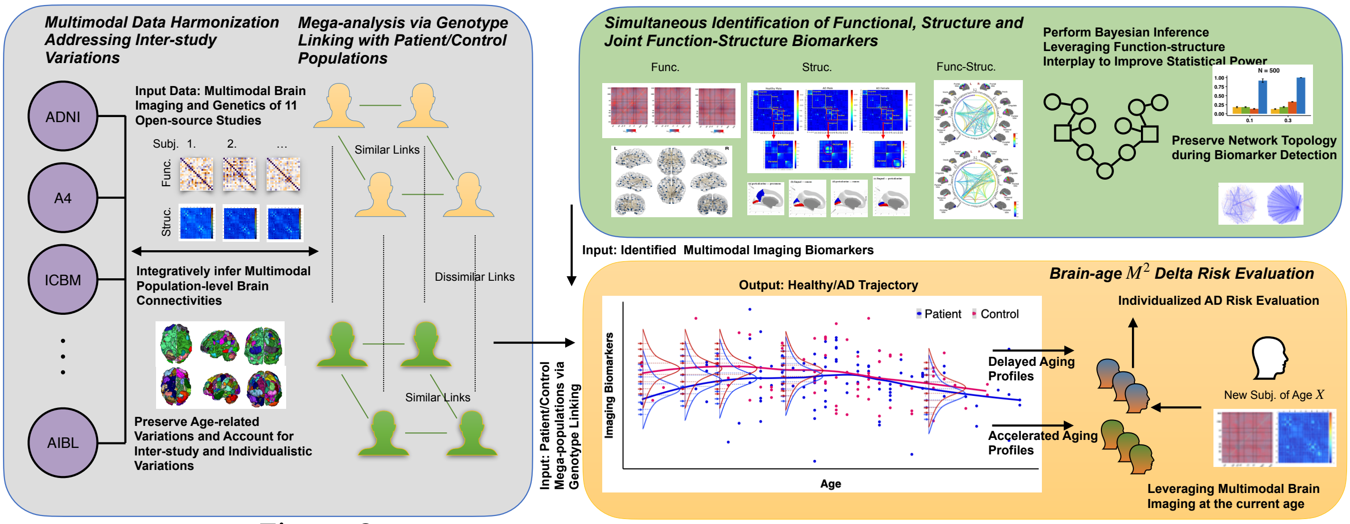

Figure 3: Schematic design of future research on neurodegeneration.

Alzheimer's Disease/Aging. For effective interventions against Alzheimer's disease (AD), while the emerging research offer extensive insights into the genetic, neural and cognitive fingerprints of the debilitating impairment, they also highlight challenges for its **early risk detection, a key to halting AD progression and even cure**. Statistical network science, a collaboration between network science and statistical theory, is well suited for constructing generative models of the brain networks, integrating multiple levels of observations and complexity and addressing the high heterogeneity, low precision and generalizability issue often found in **trajectory-based AD risk evaluation**. Yet, to date, limited statistical network methods have been developed to map the trajectories of AD versus healthy aging from earlier life stages to late adulthood integrating cognitive, functional, and anatomical dimensions, representing a major lost of opportunity to improve statistical validity and efficiency without which misinterpretation of effect can occur with inflated type I error, and/or loss of power. In Figure 3, I aim to interface advances in **statistical network science** to differentiate AD trajectories from normal aging integrating multiple cognitive, functional, and anatomical dimensions, identify joint structure-signal, function-structure signatures of healthy aging and AD and build a **predictive mechanism using multi-modality brain-age delta** targeting the earlier adulthood and pre-clinical stages.

Brains, Behaviors and Genetics. Thanks to the availability of large-scale collaborative projects such as the **Adolescent Brain Cognitive Development (ABCD) study**, methodology advancement using latent variable models for brain connectivity, behaviors and genetics are possible. The ABCD study tracks the cognitive development of around $12,000$ children from adolescent to young adulthood. Using the ABCD data, I aim to develop statistical sound methods that view the brain connectivity as networks and jointly analyze brain connectivity together with genetics and behaviors. This topic aligns with funding priorities at the National Institute of Health.

References

[1] Selena Shuo Wang, Subhadeep Paul, and Paul De Boeck. Joint latent space model for social networks with multivariate attributes. *arXiv preprint arXiv:1910.12128*, 2019.

[2] Shuo Wang. Joint analysis of social and item response networks with latent space models. Master's thesis, The Ohio State University, 2019.

[3] Selena Wang. jlsm: Joint latent space model for social networks and attributes. r package version 1. available online at: https://cran.r-project. org/package=jlsm. 2021.

[4] Selena Wang. *Modeling Dependence Between a Network and Item Responses*. PhD thesis, The Ohio State University, 2022.

[5] Selena Wang and Jared Edgerton. Resilience to stress in bipartite networks: Application to the islamic state recruitment network. *Journal of Complex Networks*, 10(4):cnac017, 2022.

[6] Selena Wang and Paul De Boeck. Understanding the role of subpopulations and reliability in between-group studies. *Behavior Research Methods*, pages 1–16, 2022.

[7] Selena Wang, Paul De Boeck, and Marcel Yotebieng. Heywood cases in unidimensional factor models and item response models for binary data. *arXiv preprint arXiv:2108.04925*, 2021.