
Subtypes of Relapsing-Remitting Multiple Sclerosis Identified by Network Analysis

Quentin Howlett-Prieto¹, Chelsea Oommen¹, Michael D. Carrithers¹, Donald C. Wunsch II², and Daniel B. Hier^{1,2,*}

¹*Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago IL USA*

²*Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla MO USA*

Correspondence*:
Daniel B. Hier
hierd@mst.edu

ABSTRACT

The objective of this study was to use network analysis to identify subtypes of relapsing-remitting multiple sclerosis subjects based on their cumulative signs and symptoms. We reviewed the electronic medical records of 120 subjects with relapsing-remitting multiple sclerosis and recorded signs and symptoms. Signs and symptoms were mapped to a neuro-ontology and then collapsed into 16 superclasses by subsumption and normalized. Bipartite (subject-feature) and unipartite (subject-subject) network graphs were created using Gephi. Degree and weighted degree were calculated for each node. Graphs were partitioned into communities using the modularity score. Feature maps were used to visualize differences in features by the community. Network analysis of the unipartite graph yielded a higher modularity score (0.49) than the bipartite graph (0.247). Network analysis can partition multiple sclerosis subjects into communities based on signs and symptoms. Communities of subjects with predominant motor, sensory, pain, fatigue, cognitive, behavior, and fatigue features were found. Larger datasets and additional partitioning algorithms are needed to confirm these results and elucidate their clinical significance.

Keywords: multiple sclerosis, phenotype, network analysis, communities, modularity

INTRODUCTION

Multiple sclerosis (MS) is one of several immune-mediated demyelinating diseases of the central nervous system that includes transverse myelitis, optic neuritis, neuromyelitis optical, acute disseminated encephalomyelitis, and acute hemorrhagic leukoencephalopathy [1]. MS has traditionally been divided into four phenotypes based on clinical course that include relapsing-remitting multiple sclerosis (RRMS), primary progressive multiple sclerosis (PPMS), secondary progressive multiple sclerosis (SPMS), and relapsing progressive multiple sclerosis (RPMS) [2]. In 2013, the criteria for MS phenotypes were revised to delete RPMS [3, 4, 5, 6]. More recently, there has been a trend to classify MS cases by activity (active or inactive) or by phase (relapsing or progressive) [4, 5, 7]. Furthermore, the clinical presentation of MS may differ by race [8, 9, 10]. Studies have suggested more cognitive

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

impairment [8], a higher prevalence of optico-spinal and transverse myelitis presentations, [9], and a faster rate of brain and retinal atrophy in African American patients with MS [10].

Another potential way of subtyping neurologic diseases is by the burden of neurological signs and symptoms. This approach has been useful in identifying subtypes of frontotemporal dementia (behavioral and aphasic variants), multiple system atrophy (cerebellar and Parkinsonian variants), amyotrophic lateral sclerosis (typical, primary lateral sclerosis, progressive bulbar atrophy, progressive muscular atrophy, and other variants), and Guillain-Barre syndrome (typical, Miller-Fisher variant, and axonal variants) [11, 12, 13, 14]. Poser [15] emphasized that MS may have a variable onset with symptoms as diverse as optic neuritis, facial pain, hemifacial spasm, Lhermitte sign, transverse myelitis, limb weakness, limb numbness, urinary retention, dysmetria, intention tremor, incoordination, dysarthria, hearing loss, color blindness, gait disturbance, and diplopia. Subtypes of multiple sclerosis based on clinical presentation (signs and symptoms) are recognized [16, 17, 18, 19]. MS patients can present with different predominant symptoms that may emphasize tremor [20], ataxia [21], visual symptoms [22, 23], sensory symptoms including numbness and paresthesias [24, 25, 26, 27], pyramidal tract findings such as weakness, hyperreflexia, spasticity, and hypertonia [28, 29, 30], or spinal cord presentations that include paraparesis, sphincter dysfunction, and sensory levels [31, 32]. Some patients present with cognitive impairment [33, 34], dysarthria [35], dysautonomia [36], depression and psychiatric disturbances [37], imbalance [38] or fatigue [39]. It is of interest whether identifiable subtypes of MS based on predominant clinical presentation (sensory, cerebellar, motor, visual, cognitive, fatigue, etc.) can be reliably identified. Clinical subtypes could differ with regard to biomarker profile, prognosis, course, and response to treatment. The network analysis of signs and symptoms could assist in identifying clinically significant subtypes of MS.

A network (also called a graph) is an assembly of nodes that are interconnected by edges [40]. When all connected nodes come from the same class, the graph is called unipartite. When each node is connected to a node of a second class, the graph is called bipartite. Networks can be created from large medical datasets and analyzed without a priori knowledge of how the nodes connect [41]. Algorithms can partition networks into communities of interconnected nodes (also called clusters) [42]. Barabási [43] defines a community as “a locally dense connected subgraph in a network (page 325)” and that “...we expect nodes that belong to a community to have a higher probability of linking to other members of that community than to nodes that do not belong to the same community...” Many partitioning algorithms depend upon the maximization of a metric called modularity which measures how well each community is separated from other communities.

Network analysis has proven useful in visualizing the complex relationships between the phenotypes, genes, proteins, and metabolic pathways that underlie human diseases (Table 1) [44, 45]. Network analysis has been applied to identifying potential genetic causes of autism [46] as well as the clustering of autism subjects by phenotype [47, 48]. Network analysis has been used to identify genes that may govern MS susceptibility [49] as well as brain areas that may undergo disconnection in MS [50, 51, 52, 53].

Proposed Approach

Our goal was to determine if clinical subtypes of RRMS based signs and symptoms could be identified. We found 244 unique neurologic signs and symptoms in a cohort of subjects with relapsing-remitting MS and collapsed them into sixteen superclasses. The count of signs and symptoms in each superclass was normalized for each subject. A bipartite graph was created in Gephi, with each subject node

Graph	Domain	node	node	edge	Citation
unipartite	phenotypes	phenotype	phenotype	similarity	[55]
unipartite	metabolism	metabolic pathway	metabolic pathway	interaction	[56]
unipartite	diseases	disease	disease	phenotype	[57]
unipartite	proteins	protein	protein	interaction	[57]
bipartite	proteins	disease	protein	disease-gene	[57]
bipartite	heart disease	heart disease	disease	co-morbidity	[58]
unipartite	autism	gene	gene	co-expression	[59]
unipartite	autism	gene	gene	co-expression	[46]
bipartite	autism	autism	disease	association	[60]
unipartite	autism	subject	subject	phenotype	[48, 47]
unipartite	brain disease	brain area	brain area	connectivity	[61]
unipartite	interactome	disease	disease	metabolism	[62]
unipartite	MS	brain area	brain area	connectivity	[50, 51, 52, 53]
unipartite	MS	gene	gene	susceptibility	[49]
unipartite	MS	MS patient	MS patient	sign/symptom	†
bipartite	MS	MS patient	phenotype	sign/symptom	†

† This paper

Table 1. Some examples of the application of network analysis to problems in human disease. This list is representative and not definitive. In general, network analysis begins with identifying nodes and edges. When nodes are connected to nodes of the same class, the graph is unipartite. When nodes are connected to a node of a second class, the graphs are bipartite. Edges connect the nodes.

connected to each of the sixteen superclass nodes by an edge proportional to the normalized count of signs and symptoms. Distances between subjects were calculated by the cosine similarity of their signs and symptoms. A unipartite graph was created in Gephi where the nodes were subjects, and the edges were subject similarities. The unipartite and bipartite graphs were partitioned into communities based on the Louvain algorithm [54]. We used modularity scores to evaluate the quality of the partition and heat maps to evaluate the composition of the communities.

METHODS

Subjects

One hundred and twenty MS subjects followed at the University of Illinois-Neuroscience Center were enrolled in the University of Illinois at Chicago (UIC) Neuroimmunology Biobank between August 2018 and March 2020 (mean age 42.7 ± 12.8 years, 73% female, 27% male, 58% Black, 42% White). The Biobank is approved by the Institutional Review Board (IRB) of the University of Illinois College of Medicine. All subjects provided informed written consent at enrollment. Subjects were between 18-80 years old and had a diagnosis of RRMS based on the 2017 McDonald criteria [63]. Subjects had been recruited for a study of blood biomarkers in MS where RRMS was an inclusion criterion, and progressive MS was an exclusion criterion. Seven subjects with normal neurological examinations were excluded from the analysis leaving a final study sample of 113 subjects.

Neuro-phenotyping

The neurological progress notes from the electronic health record of all subjects were reviewed, and neurological signs and symptoms were recorded [64]. For the purposes of this study, the cumulative signs and symptoms (both active and resolved) of each subject were recorded. Signs and symptoms were mapped to concepts in a neuro-ontology that had 1600 possible concepts [65]. Subjects had 13.2 ∓ 9.2

signs and symptoms (mean \mp standard deviation). The 113 subjects had 1453 total signs and symptoms (244 unique signs and symptoms). We used the hierarchy of the neuro-ontology [66] to collapse by subsumption the signs and symptoms into 16 superclasses that included *behavior*, *cognitive*, *cranial nerve*, *eye movement*, *fatigue*, *gait*, *hyperreflexia*, *hypertonia*, *incoordination*, *pain*, *sensory*, *speech*, *sphincter*, *tremor*, *vision*, and *weakness*. Each subject was represented as a 17-dimension vector. The first element of the vector was the case identification label, and the subsequent sixteen elements were the count of signs and symptoms for each of the sixteen superclasses. The final dataset was a 113 x 17 element array with rows as instances (subjects) and columns as features (superclasses of signs and symptoms). A normalized array was created by normalizing each feature over the interval [0,1] using the *continuize* widget in Orange 3.32.0 [67]. For the availability of python programs and datasets, see the Data Availability statement.

Network analysis, distance metrics, feature maps

Network analyses were performed on normalized 113 x 17 data arrays [67, 68, 69]. Bipartite networks were visualized in Gephi 0.9.7 using a variety of layouts, with the final analysis using the Force Atlas layout with a repulsion force of 10,000. Visual inspection showed Force Atlas to have the optimal spacing of nodes and clarity of visualization. The bipartite network contained nodes of subjects and features (signs and symptoms) as nodes with a magnitude of the edges connecting subjects to features equal to the normalized feature score for each subject. In the bipartite networks, there were no direct subject-subject or feature-feature edges. Node sizes were proportional to the average weighted degree of each node. Communities were named based on their predominant features. Nodes were colored by their community membership, and colors were used consistently across graphs based on feature predominance. Edge widths were proportional to edge weight for the bipartite graphs.

The unipartite networks were based on distances between subjects derived from the feature vectors for each subject. Distances were calculated in Orange using the *distances* widget for Pearson, Euclidean, and cosine distances. Visual inspection of the network graphs showed that the cosine-based graphs were superior to those based on the Pearson or Euclidean distances. Only the cosine distances were retained for further analysis [70]. For the unipartite graphs, all nodes were subjects, and the edges were subject similarity based on the cosine distances. Node size was proportional to degree (number of edges for each node). Edge width was fixed.

Gephi was used to partition the unipartite and bipartite networks into communities based on the Louvain algorithm [54]. The Louvain algorithm maximizes modularity (a measure of community separation). Modularity rises from 0.0 as the number of intra-community edges increases relative to inter-community edges. Larger values of modularity reflect a more robust separation of the communities. The degree, average degree, and modularity class for each node were calculated by Gephi. Modularity resolution was set to 1.0 for the normalized unipartite graph and 1.15 for the normalized bipartite graph. For the normalized unipartite graph, two subjects were excluded from the final analysis as they formed communities with only one node. For the normalized unipartite graph, a cosine distance threshold of 0.4 was used to exclude weak edges. Feature means for each community were calculated by SPSS 28 (IBM, Chicago, IL). Differences between community feature means were tested by one-way ANOVA (SPSS). Feature maps were created with the *heat map* widget from Orange.

unipartite communities	bipartite communities				
	fatigue	behavior	hypertonia/weakness	gait/sphincter	sensory
pain	0	24	6	10	21
fatigue	72	2	10	2	2
cognitive	0	0	9	9	9
sensory	0	0	2	3	56
gait/weakness/hypertonia	0	6	36	36	2

Table 2. Community concordances between bipartite and unipartite graphs as measured by Jaccard Index. Jaccard index is shown multiplied by 100. Fatigue and sensory communities show high concordance across graphs. The hypertonia/weakness and gait/sphincter communities from the bipartite graph overlap with the gait/weakness/hypertonia community from the unipartite graph.

The concordance for set membership between communities was measured by the Jaccard Index [71] where J is the Jaccard Index and A and B are the set memberships of two communities:

$$J = \frac{A \cap B}{A \cup B}.$$

RESULTS

Bipartite network analysis

The normalized feature bipartite graph was partitioned into five communities (Figure 1a) with a modularity score of 0.247. Communities were named and color-coded by the one or two features with highest community means: fatigue ($n=23$), behavior ($n=10$), hypertonia/weakness ($n=33$), gait/sphincter ($n=22$), and sensory ($n=25$) (Figure 1b). ANOVA showed significant differences between communities for behavior ($p < .001$), cranial nerve ($p = .008$), eye movements ($p < .001$), fatigue ($p < .001$), gait ($p = .029$), hyperreflexia ($p = .031$), hypertonia ($p < .001$), incoordination ($p = .002$), sensory ($p = .018$), sphincter ($p = .021$), tremor ($p = .006$), and weakness ($p = .001$).

Unipartite network analysis

The normalized features unipartite graph was partitioned into five communities (Figure 2a) with a modularity score of 0.49. Communities were named by their predominant features: Community 1 (pain), Community 2 (fatigue), Community 3 (cognitive), Community 4 (sensory), and Community 5 (weakness/gait/hypertonia) (Figure 2b). ANOVA analysis showed significant differences between communities for behavior ($p = .033$), cognitive ($p < .001$), eye movements ($p = .047$), fatigue ($p < .001$), gait ($p < .001$), hyperreflexia ($p = .014$), hypertonia ($p < .001$), incoordination ($p < .001$), pain ($p < .001$), sensory ($p < .001$), and weakness ($p = .037$).

Community overlap across graphs

Although partitioning of the the bipartite and unipartite graphs produced somewhat different communities, similarities between community membership across graphs was notable. We used the Jaccard Index (a measure of set similarity) to assess similarity between communities. Membership for the *fatigue* and *sensory* communities was similar for the unipartite and bipartite graphs (Table 2). The unipartite graph community *gait/weakness/hypertonia* showed similarity to the bipartite graph communities *hypertonia/weakness* and *gait/sphincter*.

DISCUSSION

Multiple sclerosis can present as sensory loss, weakness, incoordination, sphincter disturbance, diplopia, visual loss, cognitive impairment, fatigue, or even pain. We have used network analysis to explore whether we could identify distinct clinical subtypes of multiple sclerosis based on signs and symptoms. We first mapped the signs and symptoms of a cohort of multiple sclerosis subjects to concepts from a neuro-ontology. We then created a bipartite graph, where subjects and their signs and symptoms were nodes in a graph (Figure 1a). In a bipartite graph, subjects are connected to signs and symptoms and not to other subjects. When the signs and symptoms of a subject are converted to vectors, distances between subjects can be calculated so that subject nodes can be connected to other subjects to form a unipartite graph (Figure 2a). Network analysis allowed us to identify communities of multiple sclerosis subjects who shared signs and symptoms in common. Partitioning of the unipartite and bipartite graphs based on modularity score identified communities with strong fatigue and sensory feature predominance. Both partitions had communities characterized by weakness that was combined with hypertonia or gait disorders. Partitioning of the bipartite graph produced a small community with behavioral changes (depression, anxiety, etc.) and a gait/sphincter community. Partitioning of the unipartite graph produced a small community with cognitive disorders and a medium-sized community with pain (Figure 2b).

Partitions of the unipartite graph yielded higher modularity scores than the bipartite graph, suggesting that the partition of the unipartite graph was more robust. The partition of the unipartite graph (Figure 2a) gave the highest modularity score and showed an especially robust separation of communities. The named communities for Figure 2b (*pain, fatigue, cognitive, sensory, and gait/weakness/hypertonia*) deserve special consideration as potentially identifiable multiple sclerosis subtypes. We found a strong overlap between the fatigue and sensory communities across both graphs (Table 2 as measured by the Jaccard Index). Significant overlap between the gait/weakness/hypertonia community from the unipartite graph with the gait/sphincter and hypertonia/weakness communities from the bipartite graph was noted. Although the partitioning of networks based on features suggests that identifiable MS subtypes may exist, variability across partitions does not permit a definitive characterization of subtypes at this time.

Although we did not correlate community features with MRI findings, the communities detected may reflect the anatomic location of MS lesions [72, 73]. While some identified communities may correlate with the anatomic localization of demyelinating plaques, correlations with other disease markers such as glial fibrillary acidic protein (GFAP) and neurofilament light chain (NFL) are possible.

Two strengths of this study should be mentioned. First, community detection was done by network analysis which offers an alternative to unsupervised machine learning algorithms based on cluster analysis. Second, we used subsumption and the hierarchical organization of signs and symptoms in an ontology to reduce the number of features used in the analysis [66]. The current study demonstrates that subsumption can successfully group signs and symptoms of MS subjects into superclasses. These superclasses can be used to characterize the clinical features of communities identified by network analysis.

The current study has several limitations. The sample size was small ($N = 113$). The small sample size raises questions of selection bias which may have influenced what communities of subjects were found by network analysis. Network analysis of larger sample sizes may detect more robust communities with a different profile of predominant features. Another limitation was that we evaluated only one partitioning algorithm (Louvain). Other partitioning algorithms are available and might yield different results. We used subsumption to reduce the number of clinical features from 244 to sixteen.

Different subsumption strategies would likely yield different results. We calculated distances between subjects using the cosine distance metric; other distance metrics are available and may have resulted in different results. The modularity scores of the partitions were modest (0.25 to 0.49), suggesting that the separation between communities was modest. Another limitation was that subjects in the study were diagnosed with the RRMS phenotype. Without further analysis, our data cannot be extrapolated to other disease course phenotypes. Our analysis did not consider the race or sex of the subjects, which could influence clinical subtype [8, 9, 10]. Finally, we partitioned MS subjects based on their accumulated signs and symptoms. Examining networks based on signs and symptoms at a single point in time would be instructive.

CONCLUSIONS

MS phenotypes based on the clinical course are well-established and broadly accepted. Clinical subtypes of MS based on clinical presentation and clinical burden of signs and symptoms have long been recognized. Network analysis of signs and symptoms in MS subjects offers an avenue to identify clinical subtypes of MS. In particular, the feature maps (Figures 1b and 2b) suggest that identifiable subtypes of MS with predominant signs and symptoms related to weakness, sensory loss, behavioral changes, cognitive difficulties, pain, and fatigue deserve further investigation. Further investigation may reveal epigenetic, radiological, immunologic, or protein biomarkers that correlate with these MS clinical subtypes.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted without any commercial or financial relationships construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Design and concept by DBH and QHP. Data acquisition by MDC, DBH, QHP, and CO. Computations by DBH and QHP. Data analysis, data interpretation, and writing by DBH, QHP, MDC, CO, and DCWII. Revisions and approval by DBH, QHP, MDC, CO, and DCWII.

FUNDING

MDC acknowledges research funding from the Department of Veterans Affairs (BLR&D Merit Award BX000467) and prior support from Biogen. DCW II acknowledges research funding support from the Mary K. Finley Endowment at the Missouri University of Science and Technology.

DATA AVAILABILITY STATEMENT

Supporting data is available from the corresponding author on request.

REFERENCES

- [1] Kerr DA. The lumping and splitting of inflammatory CNS diseases. *Neurology* **66** (2006) 1466–1467.
- [2] Lublin FD, Reingold SC, et al. Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology* **46** (1996) 907–911.

- [3] Marcus JF, Waubant EL. Updates on clinically isolated syndrome and diagnostic criteria for multiple sclerosis. *The Neurohospitalist* **3** (2013) 65–80.
- [4] Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* **83** (2014) 278–286.
- [5] Lublin FD. New multiple sclerosis phenotypic classification. *European neurology* **72** (2014) 1–5.
- [6] Engelhard J, Oleske DM, Schmitting S, Wells KE, Talapala S, Barbato LM. Multiple sclerosis by phenotype in germany. *Multiple sclerosis and related disorders* **57** (2022) 103326.
- [7] Kantarci OH. Phases and phenotypes of multiple sclerosis. *CONTINUUM: Lifelong Learning in Neurology* **25** (2019) 636–654.
- [8] Weinstock-Guttman B, Jacobs L, Brownscheidle C, Baier M, Rea D, Apatoff B, et al. Multiple sclerosis characteristics in african american patients in the new york state multiple sclerosis consortium. *Multiple Sclerosis Journal* **9** (2003) 293–298.
- [9] Cree B, Khan O, Bourdette D, Goodin D, Cohen J, Marrie R, et al. Clinical characteristics of african americans vs caucasian americans with multiple sclerosis. *Neurology* **63** (2004) 2039–2045.
- [10] Caldito NG, Saidha S, Sotirchos ES, Dewey BE, Cowley NJ, Glaister J, et al. Brain and retinal atrophy in african-americans versus caucasian-americans with multiple sclerosis: a longitudinal study. *Brain* **141** (2018) 3115–3129.
- [11] Miki Y, Foti SC, Asi YT, Tsushima E, Quinn N, Ling H, et al. Improving diagnostic accuracy of multiple system atrophy: a clinicopathological study. *Brain* **142** (2019) 2813–2827.
- [12] Ghosh S, Lippa CF. Clinical subtypes of frontotemporal dementia. *American Journal of Alzheimer's Disease & Other Dementias* **30** (2015) 653–661.
- [13] Takeda T, Kitagawa K, Arai K. Phenotypic variability and its pathological basis in amyotrophic lateral sclerosis. *Neuropathology* **40** (2020) 40–56.
- [14] Dimachkie MM, Barohn RJ. Guillain-barré syndrome and variants. *Neurologic clinics* **31** (2013) 491–510.
- [15] Poser C. Onset symptoms of multiple sclerosis. *Journal of neurology, neurosurgery, and psychiatry* **58** (1995) 253.
- [16] Ford H. Clinical presentation and diagnosis of multiple sclerosis. *Clinical Medicine* **20** (2020) 380.
- [17] Loonstra FC, De Ruiter LR, Doesburg D, Lam KH, Van Lierop ZY, Moraal B, et al. Project y: The search for clues explaining phenotype variability in ms. *Multiple Sclerosis and Related Disorders* **57** (2022) 103337.
- [18] Naismith RT, Trinkaus K, Cross A. Phenotype and prognosis in african-americans with multiple sclerosis: a retrospective chart review. *Multiple Sclerosis Journal* **12** (2006) 775–781.
- [19] Cree BA, Reich DE, Khan O, De Jager PL, Nakashima I, Takahashi T, et al. Modification of multiple sclerosis phenotypes by african ancestry at hla. *Archives of neurology* **66** (2009) 226–233.
- [20] Ayache SS, Chalah MA, Al-Ani T, Farhat WH, Zouari HG, Créange A, et al. Tremor in multiple sclerosis: The intriguing role of the cerebellum. *Journal of the Neurological Sciences* **358** (2015) 351–356.
- [21] Mills RJ, Yap L, Young CA. Treatment for ataxia in multiple sclerosis. *Cochrane Database of Systematic Reviews* (2007).
- [22] Jacobs DA, Galetta SL. Multiple sclerosis and the visual system. *Ophthalmology Clinics of North America* **17** (2004) 265–73.
- [23] Costello F. Vision disturbances in multiple sclerosis. *Seminars in neurology* **36** (2016) 185–195.
- [24] Rae-Grant AD, Eckert NJ, Bartz S, Reed JF. Sensory symptoms of multiple sclerosis: a hidden reservoir of morbidity. *Multiple Sclerosis Journal* **5** (1999) 179–183.

- [25] O'Connor AB, Schwid SR, Herrmann DN, Markman JD, Dworkin RH. Pain associated with multiple sclerosis: systematic review and proposed classification. *PAIN* **137** (2008) 96–111.
- [26] Kratz AL, Whibley D, Alschuler KN, Ehde DM, Williams DA, Clauw DJ, et al. Characterizing chronic pain phenotypes in multiple sclerosis: a nationwide survey study. *Pain* **162** (2021) 1426.
- [27] Sanders E, Arts R. Paraesthesiae in multiple sclerosis. *Journal of the neurological sciences* **74** (1986) 297–305.
- [28] Rizzo M, Hadjimichael O, Preiningerova J, Vollmer T. Prevalence and treatment of spasticity reported by multiple sclerosis patients. *Multiple Sclerosis Journal* **10** (2004) 589–595.
- [29] Hoang PD, Gandevia SC, Herbert RD. Prevalence of joint contractures and muscle weakness in people with multiple sclerosis. *Disability and rehabilitation* **36** (2014) 1588–1593.
- [30] Cordani C, Hidalgo de la Cruz M, Meani A, Valsasina P, Esposito F, Pagani E, et al. Mri correlates of clinical disability and hand-motor performance in multiple sclerosis phenotypes. *Multiple Sclerosis Journal* **27** (2021) 1205–1221.
- [31] Nociti V, Cianfoni A, Mirabella M, Caggiula M, Frisullo G, Patanella AK, et al. Clinical characteristics, course and prognosis of spinal multiple sclerosis. *Spinal Cord* **43** (2005) 731–734.
- [32] Wiesel PH, Norton C, Glickman S, Kamm MA. Pathophysiology and management of bowel dysfunction in multiple sclerosis. *European journal of gastroenterology & hepatology* **13** (2001) 441–448.
- [33] Amato MP, Zipoli V, Portaccio E. Cognitive changes in multiple sclerosis. *Expert review of neurotherapeutics* **8** (2008) 1585–1596.
- [34] De Meo E, Portaccio E, Giorgio A, Ruano L, Goretti B, Niccolai C, et al. Identifying the distinct cognitive phenotypes in multiple sclerosis. *JAMA neurology* **78** (2021) 414–425.
- [35] Noffs G, Perera T, Kolbe SC, Shanahan CJ, Boonstra FM, Evans A, et al. What speech can tell us: A systematic review of dysarthria characteristics in multiple sclerosis. *Autoimmunity reviews* **17** (2018) 1202–1209.
- [36] Miglis MG, Muppidi S. Autonomic dysfunction in multiple sclerosis and other updates on recent autonomic research. *Clinical Autonomic Research* **28** (2018) 391–393.
- [37] Sá MJ. Psychological aspects of multiple sclerosis. *Clinical neurology and neurosurgery* **110** (2008) 868–877.
- [38] Soyuer F, Mirza M, Erkorkmaz Ü. Balance performance in three forms of multiple sclerosis. *Neurological research* **28** (2006) 555–562.
- [39] Induruwa I, Constantinescu CS, Gran B. Fatigue in multiple sclerosis—a brief review. *Journal of the neurological sciences* **323** (2012) 9–15.
- [40] Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers* **1** (2021) 1–18.
- [41] Kramer J, Boone L, Clifford T, Bruce J, Matta J. Analysis of medical data using community detection on inferred networks. *IEEE Journal of Biomedical and Health Informatics* **24** (2020) 3136–3143.
- [42] Newman ME. Modularity and community structure in networks. *Proceedings of the national academy of sciences* **103** (2006) 8577–8582.
- [43] Barabási AL. *Network Science* (Cambridge University Press: Cambridge UK) (2016).
- [44] Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nature reviews genetics* **12** (2011) 56–68.
- [45] Carter H, Hofree M, Ideker T. Genotype to phenotype via network analysis. *Current opinion in genetics & development* **23** (2013) 611–621.

- [46] Emberti Gialloreti L, Enea R, Di Micco V, Di Giovanni D, Curatolo P. Clustering analysis supports the detection of biological processes related to autism spectrum disorder. *Genes* **11** (2020) 1476.
- [47] Matta J, Zhao J, Ercal G, Obafemi-Ajayi T. Applications of node-based resilience graph theoretic framework to clustering autism spectrum disorders phenotypes. *Applied network science* **3** (2018) 1–22.
- [48] Matta J, Dobrino D, Yeboah D, Howard S, Yasser EM, Obafemi-Ajayi T. Connecting phenotype to genotype: Phewas-inspired analysis of autism spectrum disorder. *Frontiers in Human Neuroscience* **16** (2022).
- [49] Baranzini SE, Khankhanian P, Patsopoulos NA, Li M, Stankovich J, Cotsapas C, et al. Network-based multiple sclerosis pathway analysis with gwas data from 15,000 cases and 30,000 controls. *The American Journal of Human Genetics* **92** (2013) 854–865.
- [50] Jandric D, Lipp I, Paling D, Rog D, Castellazzi G, Haroon H, et al. Mechanisms of network changes in cognitive impairment in multiple sclerosis. *Neurology* **97** (2021) e1886–e1897.
- [51] Schiavi S, Azzari A, Mensi A, Graziano N, Daducci A, Bicego M, et al. Classification of multiple sclerosis patients based on structural disconnection: A robust feature selection approach. *Journal of Neuroimaging* (2022).
- [52] Schoonheim MM, Meijer KA, Geurts JJ. Network collapse and cognitive impairment in multiple sclerosis. *Frontiers in neurology* **6** (2015) 82.
- [53] Fleischer V, Radetz A, Ciolac D, Muthuraman M, Gonzalez-Escamilla G, Zipp F, et al. Graph theoretical framework of brain networks in multiple sclerosis: a review of concepts. *Neuroscience* **403** (2019) 35–53.
- [54] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008** (2008) P10008.
- [55] Van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA. A text-mining analysis of the human phenome. *European journal of human genetics* **14** (2006) 535–542.
- [56] Seo D, Lee MH, Yu SJ. Development of network analysis and visualization system for kegg pathways. *Symmetry* **7** (2015) 1275–1288.
- [57] Vanunu O, Mager O, Ruppin E, Shlomi T, Sharan R. Associating genes and protein complexes with disease via network propagation. *PLoS computational biology* **6** (2010) e1000641.
- [58] Zhou D, Wang L, Ding S, Shen M, Qiu H. Phenotypic disease network analysis to identify comorbidity patterns in hospitalized patients with ischemic heart disease using large-scale administrative data. *Healthcare* **10** (2022) 80.
- [59] Alonso-Gonzalez A, Rodriguez-Fontenla C, Carracedo A. De novo mutations (dnms) in autism spectrum disorder (asd): Pathway and network analysis. *Frontiers in genetics* **9** (2018) 406.
- [60] Li X, Liu G, Chen W, Bi Z, Liang H. Network analysis of autistic disease comorbidities in chinese children based on icd-10 codes. *BMC Medical Informatics and Decision Making* **20** (2020) 1–11.
- [61] Stam CJ. Modern network science of neurological disorders. *Nature Reviews Neuroscience* **15** (2014) 683–695.
- [62] Lee DS, Park J, Kay K, Christakis NA, Oltvai ZN, Barabási AL. The implications of human metabolic network topology for disease comorbidity. *Proceedings of the National Academy of Sciences* **105** (2008) 9880–9885.
- [63] Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetsee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the mcdonald criteria. *The Lancet Neurology* **17** (2018) 162–173.

- [64] Hier DB, Yelugam R, Azizi S, Wunsch III DC. A focused review of deep phenotyping with examples from neurology. *European Scientific Journal* <https://doi.org/10.19044/esj.2022.v18n4p4> (2022) 4–19.
- [65] Hier DB, Brint SU. A neuro-ontology for the neurological examination. *BMC Medical Informatics and Decision Making* **20** (2020) 1–9.
- [66] Wunsch III DC, Hier DB. Subsumption is a novel feature reduction strategy for high dimensionality datasets. *European Scientific Journal* **18** (2022) 20–36. Accessed August 12, 2022.
- [67] Demsar J, Zupan B. Orange: Data mining fruitful and fun-a historical perspective. *Informatica* **37** (2013) 55.
- [68] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *Proceedings of the international AAAI conference on web and social media* **3** (2009) 361–362.
- [69] Hagberg A, Swart P, S Chult D. Exploring network structure, dynamics, and function using networkx. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States) (2008).
- [70] Hier DB, Kopel J, Brint SU, Wunsch DC, Olbricht GR, Azizi S, et al. Evaluation of standard and semantically-augmented distance metrics for neurology patients. *BMC medical informatics and decision making* **20** (2020) 1–15.
- [71] JACCARD P. Distribution de la flore alpine dans le bassin des drouces et dans quelques regions voisines. *Bull Soc Vaud Sci Nat* **37** (1901) 241–272.
- [72] Huber SJ, Paulson GW, Chakeres D, Pakalnis A, Brogan M, Phillips BL, et al. Magnetic resonance imaging and clinical correlations in multiple sclerosis. *Journal of the neurological sciences* **86** (1988) 1–12.
- [73] Stevens JC, Farlow MR, Edwards MK, Yu Pl. Magnetic resonance imaging: clinical correlation in 64 patients with multiple sclerosis. *Archives of neurology* **43** (1986) 1145–1148.

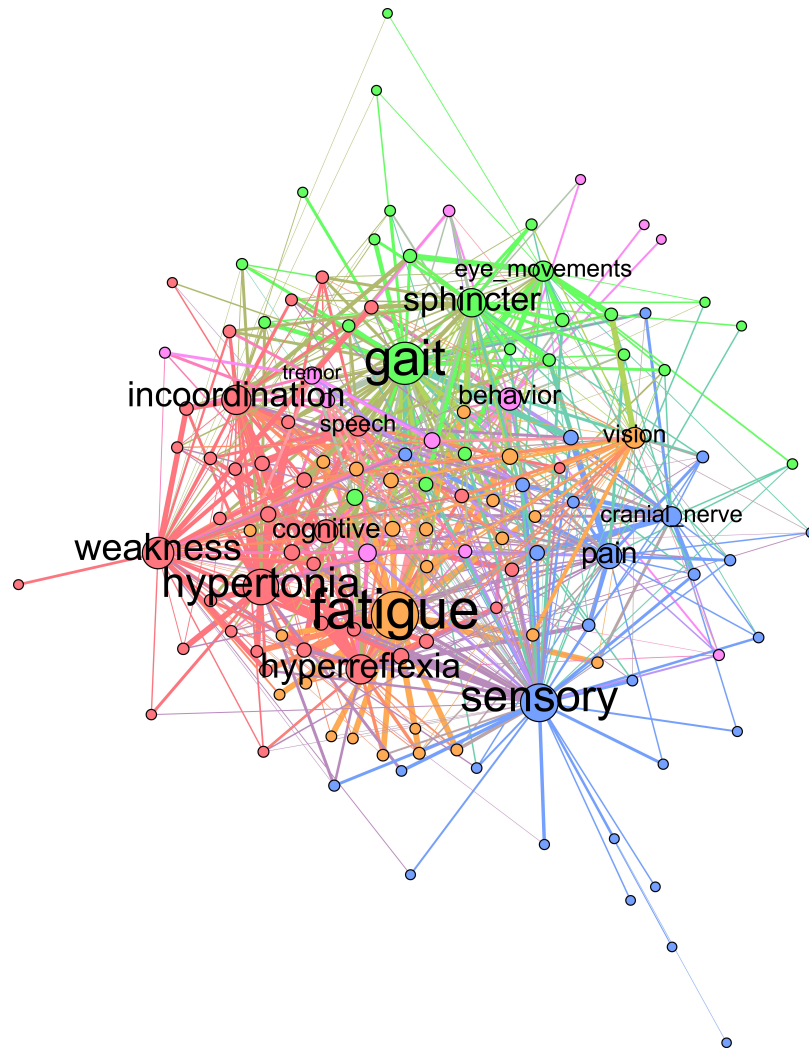


Figure 1a. Bipartite graph of normalized features. Labeled nodes are features; unlabeled nodes are subjects. Weighted connections (edges) are between features and subjects.

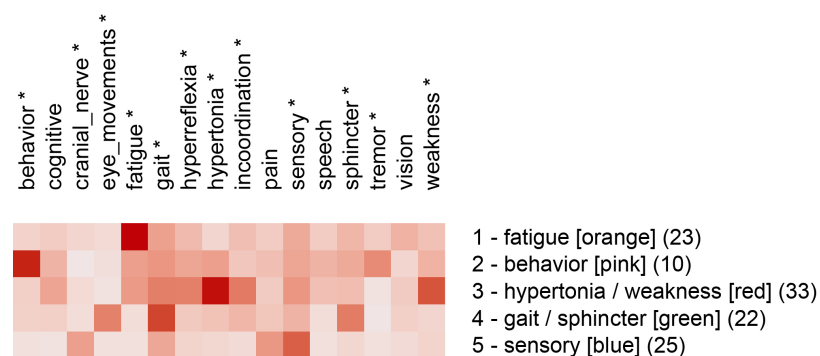


Figure 1b. Feature map of the five communities identified by partitioning the bipartite graph with unnormalized features. Asterisks indicate features that differed significantly by community (One-way ANOVA, $p < 0.05$, $df = 4$)

Figure 2. Modularity analysis of the bipartite graph of normalized data showed five communities. Community 1 was predominantly fatigue, Community 2 was predominantly behavioral, Community 3 was weakness and hypertonia, Community 4 was gait and sphincter, and Community 5 was predominantly sensory features.

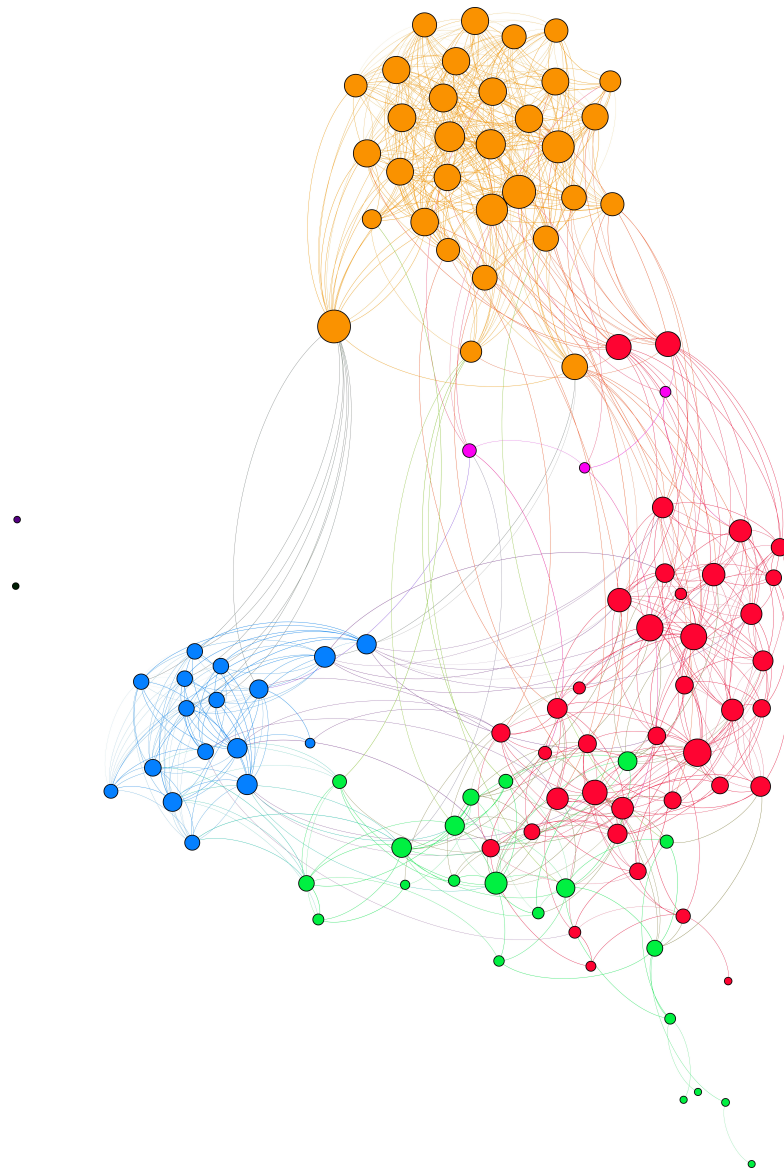


Figure 2a. Unipartite graph based on normalized features. Nodes are subjects, and node size is proportional to the number of edges. The largest communities are gait/weakness/hypertonia (red, n=38) and fatigue (orange, n=32). Note the small cognitive community (pink, n=3).

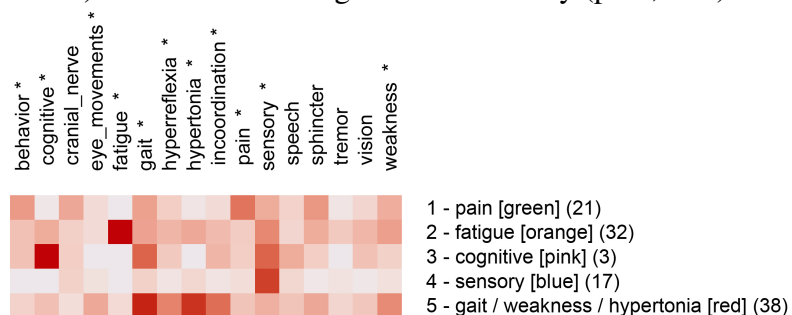


Figure 2b. Feature map of the five communities identified by partitioning the unipartite network graph based on normalized features. Asterisks indicate features that differed significantly by community (One-way ANOVA, $p < 0.05$, $df = 4$)