Clinical Psychology Review 85 (2021) 102000

Contents lists available at ScienceDirect

Clinical Psychology Review

journal homepage: www.elsevier.com/locate/clinpsychrev

Review

Networks of major depressive disorder: A systematic review

Matteo Malgaroli [a,*,1], Adam Calderon [b,1], George A. Bonanno [b]

[a] Department of Psychiatry, NYU Grossman School of Medicine, New York, NY, United States of America
[b] Department of Clinical Psychology, Teachers College, Columbia University, New York, NY, United States of America

ARTICLE INFO

Keywords:
Depression
Network analysis
Heterogeneity
Systematic review

ABSTRACT

There has been a marked increase of network studies of Major Depressive Disorder (MDD). Despite rapidly growing contributions, their findings have yet to be systematically aggregated and examined. We therefore conducted a systematic review of depression network studies using PRISMA guidelines. A total of 254 clinical and population studies were collected from ISI's Web of Science and PsycINFO, between January 2010 to May 2020. A total of 23 between-subject studies were included for review, resulting in 58 cross-sectional networks. To determine their most critical symptoms and their connections, we analyzed strength centrality rankings, and aggregated the most robust symptoms connections into a summary network. Results indicated substantial variability between study samples, depression measures, and network features. Fatigue and Depressed Mood were the most central symptoms, while Weight changes tended to have the weakest centrality. Depressed Mood and Fatigue formed two separated symptoms communities characterized by recurrent connections, with Mood-Anhedonia as the most frequent edge of MDD. Network analysis informed our understanding of MDD, suggesting the critical role of Fatigue and Depressed Mood. The study's findings are discussed in their clinical and methodological implications, including future directions for network studies of MDD.

1. Introduction

Major Depressive Disorder (MDD) is the leading cause of disability worldwide, affecting more than 300 million individuals (World Health Organization, 2017). It is among the most frequent psychiatric disorders (Kessler, Chiu, Demler, & Walters, 2005), with an estimated lifetime prevalence of 16.2% (Kessler et al., 2003). Yet, despite its pervasiveness, the diagnosis and measurement of MDD have suffered from a concerning level of variability and heterogeneity. Over 280 scales have been developed to measure MDD (Santor, Gregus, & Welch, 2006), which are often used interchangeably despite weak correlations and, notably, differences in symptom content (Fried, 2017). Diagnostic manuals such as the DSM-5 (APA, 2013) have established a lingua franca to label different depression presentations, yet MDD criteria have been shown to produce 1030 different symptom patterns (Fried & Nesse, 2015a). In part because of this high level of heterogeneity, MDD has one of the lowest inter-rater reliabilities among DSM-5 disorders (Regier et al., 2013). Individuals who share the same MDD diagnosis may endorse different symptoms, and yet be offered the same blanket interventions. As such, a considerable amount of research has been engendered to construct a comprehensive taxonomic framework of depression, with the ultimate goal of improving targeted treatment.

Under the current DSM-5 diagnostic criteria, MDD is characterized by nine symptoms: 1) Depressed mood (Mood); 2) Loss of interest or pleasure (Anhedonia); 3) Loss or increase in appetite or weight (Weight); 4) Insomnia or hypersomnia (Sleep); 5) Psychomotor agitation or retardation (Psychomotor); 6) Loss of energy or fatigue (Fatigue); 7) Feelings of worthlessness or inappropriate guilt (Worthlessness); 8) Impaired concentration or indecisiveness (Concentration); 9) Suicidal ideation, plans, or thoughts of death (Suicidal thoughts and behaviors). The diagnostic threshold of five or more symptoms results in 277 combinations meeting DSM-5 diagnostic criteria for MDD (Zimmerman, Ellison, Young, Chelminski, & Dalrymple, 2015), with the added complexity of polar-opposite presentations combined in the same symptom category (e.g., insomnia/hypersomnia). Moreover, representing MDD as a linear sum of symptoms implies that symptoms emerge homogenously from an underlying depressive disorder (Fried & Nesse, 2015b). This would suggest that all symptoms are equally important and that the individual symptoms are not relevant (Fried, 2015). Such assumptions are inconsistent with both statistical consideration (i.e., local

* Corresponding author at: One Park Avenue, 8th Floor, New York, NY 10016, United States of America.
E-mail address: matteo.malgaroli@nyulangone.org (M. Malgaroli).
[1] These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.cpr.2021.102000
Received 16 July 2020; Received in revised form 6 January 2021; Accepted 23 February 2021
Available online 1 March 2021

independence) and clinical observations of how symptoms interact (e.g., Sleep problems leading to Concentration difficulties).

In an effort to move beyond these problems, an increasingly popular approach has been to use network theory to probe the heterogeneity and etiological underpinnings of MDD (Borsboom & Cramer, 2013). Within this approach, MDD is conceptualized as an interconnected network emerging from symptoms interactions; in other words, the symptoms themselves *constitute* MDD (Borsboom, 2017; Jones, Heeren, & McNally, 2017). Symptoms are defined as distinct causal agents that, when unabated, foster the development of other symptomsbeyond a critical threshold into a new harmful equilibrium that we define as MDD (Robinaugh, Hoekstra, Toner, & Borsboom, 2019).

Thanks to the accessibility of network analytic tools (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012), numerous studies have now used the network approach to analyze clinical and population samples. Several informative overviews on network psychometrics have been published (Borsboom et al., 2016; Borsboom & Cramer, 2013; Fried et al., 2017; Fried & Cramer, 2017; McNally, 2016), as well as three systematic literature reviews (Birkeland, Greene, & Spiller, 2020; Contreras, Nieto, Valiente, Espinosa, & Vazquez, 2019; Robinaugh et al., 2019). Despite these seminal contributions, the vast corpus of network analyses results on MDD has yet to be systematically integrated and consolidated. Completing these network comparisons is essential to replicability (Borsboom et al., 2018; Forbes, Wright, Markon, & Krueger, 2017). Given Contreras et al. (2019) call to action for researchers to begin developing systematic reviews on specific psychopathologies, coupled with Robinaugh et al.'s (2019) recommendation for constructing reviews specific to network studies, the need for integrated findings from MDD network studies is imperative.

To this end, in the current review we systematically examined published network analyses on MDD intending to determine its most central and crucial symptoms to emerge across studies. More specifically, we investigated the symptom characteristics of published clinical and population-based cross-sectional network studies of MDD, specifying 1) the characteristics of MDD network studies and 2) the most recurrent centrality and edge weight indices of networks involving MDD symptoms. To our knowledge, this is the first systematic review of the literature on MDD networks.

2. Methods

2.1. Search protocol and eligibility

Studies were systematically searched, screened, and selected for inclusion following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher, Liberati, Tetzlaff, &

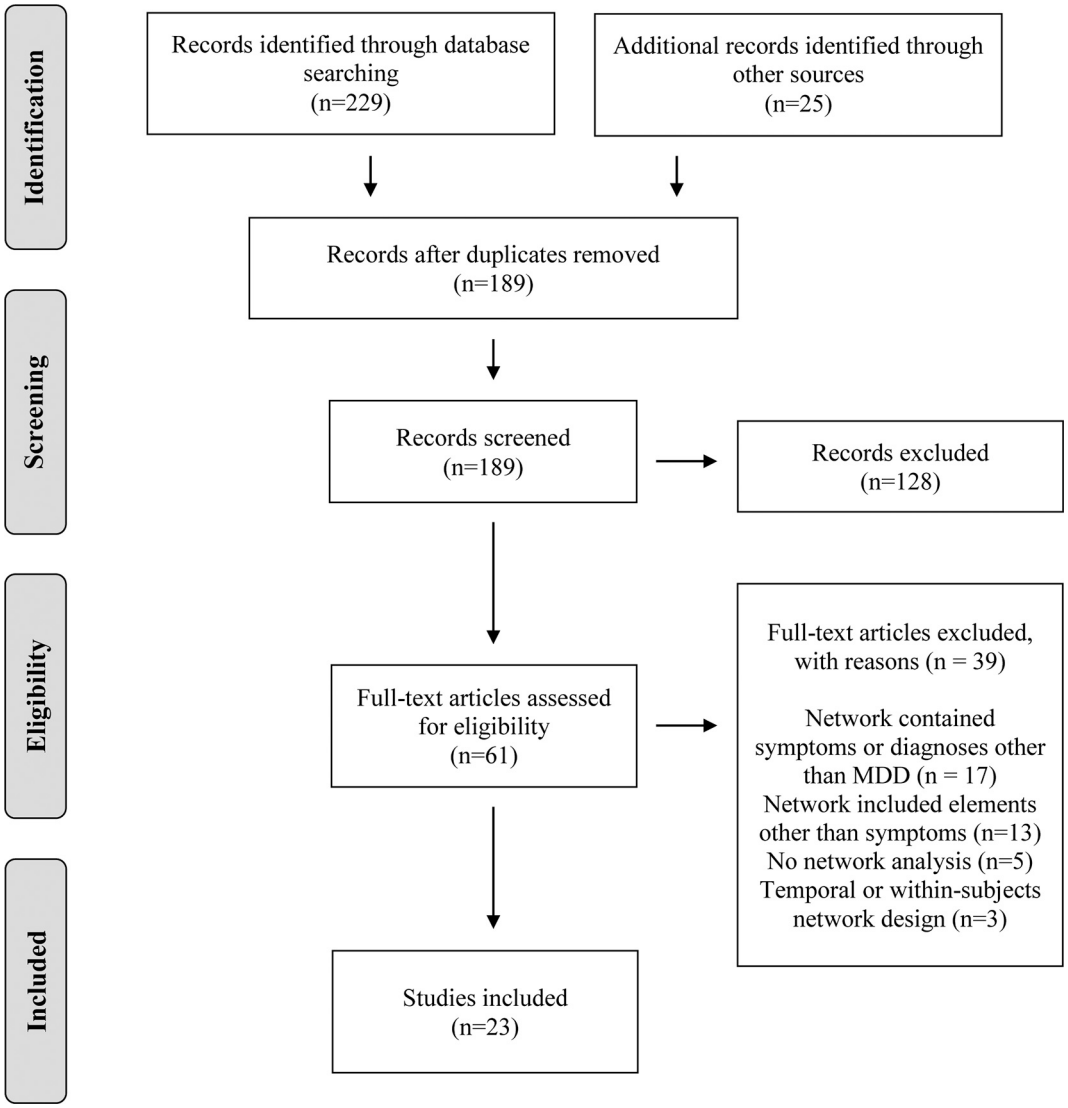

Fig. 1. PRISMA flow diagram.

Altman, 2010). Articles were found by searching PsycINFO and the Institute for Scientific Information (ISI) Web of Science. Searches were restricted to studies published between January 2010 to May 2020. The following Boolean search terms were used: (network analysis OR network approach OR network model OR network structure OR network modeling OR network theory) AND (major depressive disorder OR major depression OR MDD OR depression symptoms OR depressive symptoms OR MDD symptoms). The screening process began with each article being assessed by title and abstract and was then evaluated using the exclusion and inclusion criteria. Fig. 1 reports the PRISMA flow diagram corresponding to the search protocol, inclusion n and exclusion criteria, screening, eligibility, and final selection process.

Articles were considered eligible for further review if they met the following inclusion criteria: (1) original empirical study; (2) written in English; (3) peer-reviewed; (4) measures MDD symptoms; (5) conducted network analysis of MDD symptoms; (6) estimated a network based on cross-sectional data; (7) published between January 2010 to May 2020. Articles were excluded if: (1) commentary or editorial article; (2) book chapter; (3) review article; (4) unpublished article; (5) no network analysis; (6) no usable network model. Exclusion criteria for networks models consisted of: (6.1) used social networks; (6.2) used brain networks; (6.3) used temporal or within-subjects network design; (6.4) network included elements other than symptoms (e.g., experimental variables); (6.5) network contained symptoms or diagnoses other than depression (e.g., OCD symptoms). If a study had no remaining eligible networks, it was then excluded from the review.

2.2. Network studies characteristics

Network models are constituted by nodes (i.e. the symptoms) and edges (i.e., magnitude of their connections). We analyzed and compared the characteristics of each study and their networks along six sets of characteristics. Multiple networks from the same study were examined if present (e.g., comparisons across groups). For each set, we provide a brief description below of the most important attributes.

2.2.1. Sample characteristics

Type: If the sample was recruited from either clinical or population-based settings. Age: Mean age and/or age range of the sample. Gender: distribution within the study sample.

2.2.2. MDD symptoms

DSM edition: Which MDD diagnostic algorithm was used. *MDD Measure*: if the study assessed MDD symptoms using a self-report measure and/or structured interview. *Depression Symptoms*: The total number of symptoms of depression used in each study. The number includes additional elements of MDD as identified by the assessment instrument (Fried, 2017), and any auxiliary MDD symptoms if measured.

2.2.3. Network characteristics

Number of Models: Total number of cross-sectional MDD network models estimated in the reviewed study. *Centrality & Control*: The importance of a node (or symptom) in a network is indicated by its centrality. For psychological networks, Strength is the most evaluated centrality measure, representing the absolute sum of a symptom's connections with all other nodes. Other centrality measures include Closeness, Betweenness, and Expected Influence (Robinaugh, Millner, & McNally, 2016). Additional metrics of a node's control over the network include Predictability (Haslbeck & Waldorp, 2018), the Participation Coefficient (Letina, Blanken, Deserno, & Borsboom, 2019) and Eigenvector Centrality (Bonacich, 2007). *Symptoms Connectivity:* Global strength determines overall symptoms connectivity. This metric can be measured through statistical methods such as the Network Comparison Test (NCT; van Borkulo et al., 2017), allowing to assess global strength differences between two networks. For networks that estimated global connectivity, we report if global strength changes were associated with changes in depression levels. Clusters of locally strongly interconnected nodes were instead identified using community detection algorithms (e.g., spinglass). *Group Differences*: Describes differences in network structure between groups (if assessed), and the comparison results.

2.2.4. Network estimation

Network Parameters: More estimated parameters signify more computationally demanding networks. Total parameters were calculated as per Fried and Cramer (2017), who suggested a 3+ subjects-to-parameter ratio. *Correlation*: The correlation technique used to estimate how much symptoms jointly change (i.e., covariance matrix). This includes polychoric, tetrachoric or Spearman correlations for respectively categorical, binary, and continuous data. The resulting covariance matrix is used as input for the analytic model. *Analytic Model*: the network model used to estimate the symptoms network. Gaussian Graphical Models (GGM) are used for normally distributed data (Epskamp, Kruis, & Marsman, 2017), Association networks and Ising Models for binary data (Haslbeck, Ryan, Robinaugh, Waldorp, & Borsboom, 2019), and Mixed Graphical Models (MGM) for mixed categorical and continuous data (Haslbeck & Waldorp, 2020). Fused Graphical Lasso (FGL) models are applied when network comparisons are needed (Costantini et al., 2019).

2.2.5. Network stability

CS Coefficient: A measure of centrality stability, the Correlation Stability (CS) Coefficient is the proportion of the sample that can be dropped while maintaining a correlation of 0.70+ to the original centrality score. CS coefficients should not fall below 0.25, with 0.5+ indicating relatively stable networks (Epskamp, Waldorp, Mõttus, & Borsboom, 2018). *Edge Accuracy*: Without assessing their edges stability, portrayed in an MDD network model, the results might not be relevant, heeding caution about their interpretation (Epskamp, Waldorp, et al., 2018). Bootstrapped 95% confidence intervals estimate the accuracy of symptoms' connections (Epskamp, Waldorp, et al., 2018). Additionally, bootstrapped edge difference tests determine if these edge-weights significantly differ from one another (Epskamp, Waldorp, et al., 2018). For each network, we report the number of edges that are significantly different than at least two-thirds of the total edges in a network.

2.2.6. Reproducibility

Open Data: If a study shared their full data, the covariance matrices used to generate the networks, or their analytic code. *Open Access*: If the manuscript was openly available and free of charge as Gold open access (i.e., articles published in an open-access peer-reviewed journal) or Green open access (i.e., authors established a free, unrestrictive, online repository to provide access to their article).

2.3. Analysis of MDD networks nodes and edges

After investigating the characteristics of each study, we further examined the specific features of their networks. Node strength centrality and edge weights were collected from the cross-sectional networks in each study. If studies contained multiple networks, we included in the analysis networks from different samples and cross-sectional "temporal" snapshots (e.g., different treatment phases), while excluding models that reexamined a subsect sample from another included network (e.g., subgroup comparisons).

2.3.1. Symptoms Strength Centrality Rankings

Strength centrality rankings were examined to determine the most important MDD symptoms across studies. Higher centrality indicates symptoms with stronger associations with other symptoms in the network, which in turn is associated with driving stronger changes in other nodes over time (Robinaugh et al., 2016). The analysis focused on Strength Centrality, given the poor stability of Betweenness and Closeness in symptoms networks (Bringmann et al., 2019), and the still

nascent use of expected influence. We examined relative centrality propensities of MDD symptoms, comparing the tendencies of symptoms to be among the more or less important across the examined studies. More specifically, Strength Centrality ranking information (i.e., most to least important symptom in the network) for MDD symptoms was collected in frequency tables from each study estimating the measure. As not all studies contained all symptoms, rankings were then min-max normalized and ordered based on their median centrality rank, to be then visualized using the R package ggplot2 (Wickham, 2016).

After determining the most central symptoms across networks, differences in symptom centrality across studies was examined using logistic regression in R. The analyzed study characteristics consisted of sample size, if the network is a second observation of a baseline network, and if the study was clinical or epidemiological in nature.

2.3.2. Robust edges of MDD symptoms

To determine the most interconnected symptoms, we examined recurring *robust* edges across studies. Edges represent the level of statistical association between two symptoms, which psychological network theory interprets as patterns of mutual activations through feedback loops (Borsboom & Cramer, 2013). However, in a network model not all edges are reliably different from each other, as the estimated network model only approximates the true network. Bootstrapped difference testing was designed as a solution to identify which edges are significantly different from one another. For this analysis, we therefore gathered data exclusively from studies that presented edge-weight difference tests. We only included *robust* significant edges, defined as being significantly different from at least two-thirds of the total edges in the network in bootstrapped comparison tests (Birkeland et al., 2020). We also limited our analysis to positive robust connections, to study activation patterns between MDD symptoms.

Robust symptoms connections were then used to create a summary network graph of the most frequent edges. In the summary network edge list, the weight of a connection was determined as the proportion of that edge appearing robust across the studies. More specifically, the edge weight between two depression symptoms (e.g., Mood-Weight) was defined as the number of times said edge appeared robust, divided by the number of networks containing both symptoms (e.g. number of network from studies assessing both Mood and Weight), adjusted by the total number of reviewed networks. If separate nodes measuring the same MDD symptom existed in a network (e.g. Psychomotor Activation and Retardation), robustness was determined if one or both had robust connections with another symptom.

After determining the network of robust edges among MDD symptoms, we performed exploratory analyses to assess configurations of more deeply interconnected symptoms. In network science, community detection is the approach used to identify network subsets or partitions, which can potentially exhibit complex systems underlying the graph configuration (Lancichinetti & Fortunato, 2009). Accordingly, we performed community detection to examine the structure of the robust edges networks using the spin glass algorithm (Reichardt & Bornholdt, 2006). The analysis was conducted with the R package igraph (Csardi & Nepusz, 2006), with default parameters (i.e, 25 spins, starting and stopping temperatures respectively 1 and 0.01), running the algorithm 1000 times with a different random seed for each run, and then aggregating results on the modal cluster configuration.

The resulting summary network was visualized using the package qgraph (Epskamp et al., 2012). Each edge in the summary network was labeled with its proportion of robust edges, to further clarify connections. Edges not robust in any of the examined studies were not visualized. The Fruchterman–Reingold algorithm determined the position of each Depression node in the graph, based on the frequency and weight of symptoms connections. In the resulting graph, stronger edges represent more frequent connections between MDD symptoms across the examined studies.

3. Results

A total of 254 articles were reviewed, of which 229 were identified through database searches, and 25 by examining the reference lists of past literature reviews. After removing 65 duplicate articles, 189 articles were left for abstract review, of which 128 were removed based on exclusion criteria. Of the 61 articles left for full-text review, 39 were removed due to: all of their networks containing symptoms or diagnoses other than MDD (n = 17), including elements other than symptoms (n = 13), no presence of network analysis (n = 5), or presenting a temporal or within-subjects network design (n = 3). Supplementary materials report a list of every study excluded, along with reasons for removal (see Supplementary Materials, Table S1). The final sample consisted of 23 studies, from which 58 networks of MDD were included within the review sample. Table 1 reports their full characteristics.

3.1. Network study characteristics

3.1.1. Sample characteristics

Study samples were recruited from clinical trials and patients formally diagnosed with MDD (n = 13 studies; 30 networks), or derived from a community sample (n = 10 studies; 28 networks). In terms of data sources, studies often obtained data from national cohort studies (n = 11). In particular, the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) (n = 2; Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016; Madhoo & Levine, 2016); China, Oxford, and VCU Experimental Research on Genetic Epidemiology (CONVERGE) (n = 2; van Loo et al., 2018; Kendler, Aggen, Flint, Borsboom, & Fried, 2018); and the Netherlands Study of Depression and Anxiety (NESDA) (n = 2; Boschloo, van Borkulo, Borsboom, & Schoevers, 2016; van Borkulo et al., 2015). Fourteen studies used an adult sample (n = 29 networks), while five focused on adolescents (n = 15 networks; McElroy, Napoleone, Wolpert, & Patalay, 2019; Mullarkey, Marchetti, & Beevers, 2019; Osborn, Campbell, Ndetei, & Weisz, 2020; Schweren, van Borkulo, Fried, & Goodyer, 2018; Wasil, Venturo-Conerly, Shinde, Patel, & Jones, 2020) and four consisted of only elderly participants (n = 14 networks; Airaksinen, Gluschkoff, Kivimäki, & Jokela, 2020; An et al., 2019; de la Torre-Luque et al., 2020; Murri, Amore, Respino, & Alexopoulos, 2018). Sample size ranged from 151 to 13,035, with the average number of participants being 2713. From the 20 studies that reported age data, the cumulative mean age of participants was 42.35. Of the 23 studies that reported gender, the percentage of females ranged from 47.5 (Wasil, Venturo-Conerly, Shinde, Patel, & Jones, 2020) to 100% (Kendler et al., 2018; Santos Jr, Fried, Asafu-Adjei, & Ruiz, 2017; van Loo et al., 2018), with an average of 64.88%.

3.1.2. MDD symptoms

The DSM-IV MDD criteria were the most used (n = 15 studies; 34 networks). The most frequent self-report MDD measure was the Patient Health Questionnaire (PHQ-9; Kroenke & Spitzer, 2002), appearing in four studies. The Inventory of Depressive Symptomatology Self Report (IDS-SR; Rush et al., 1986), Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 1996), and the Composite International Diagnostic Interview (CIDI) (Kessler & Üstün, 2004) were respectively found in three studies. The median number of symptoms per network was 12 (ranging from 8 to 28), with twelve studies examining additional elements of depression (16 networks). These additional items elements primarily originated from questions of the assessment measures, such as the BDI-II (Bos et al., 2018; Hakulinen et al., 2020; Park & Kim, 2020) or CIDI (de la Torre-Luque et al., 2020; Kendler et al., 2018), with two studies including additional symptoms by design (Kendler et al., 2018; van Loo et al., 2018).

3.1.3. Network characteristics

Studies most frequently estimated either one (n = 7; Boschloo et al., 2016; Kendler et al., 2018; McWilliams, Sarty, Kowal, & Wilson, 2017;

Table 1
Summary of manuscripts included in the systematic review of cross-sectional network studies of depressive symptoms.

Reference	Sample characteristics				MDD symptoms			Network analysis									Reproducibility	
								Characteristics				Estimation			Stability			
	N	Study design	Age M (range)	Gender % female	Measure	Diagnostic manual	N	Model N[1]	Centrality & control	Symptoms connectivity[2]	Structure differences[3]	Parameters N	Correlation technique	Analytic model	CS coefficient	Edge accuracy[4]	Open Data	Open Access
Airaksinen et al. (2020)	670[a]	Community	67.37 (50+)	52%	CES-D	DSM-IV	8	2	Strength, Bet., Clo.	↑MDD; ↓GS[a]	–	36	Spearman	Ising	–	–	No	Gold
An et al. (2019)	776	Clinical	73.87 (60+)	78.6%	MADRS	DSM-5	10	5	Strength, Bet., Clo.	↑MDD; ↑GS	Similar: low/ high anxiety	55	Polychoric	GGM	–	95% CI	No	Gold
Berlim et al. (2020)	151	Clinical	40.18 (18–65)	58.3%	IDS-SR	DSM-IV	9	6	E.I., Pred.	↓MDD; ↑GS	Different: before/after SSRI treatment	45	Polychoric	GGM	0.25, 0.60[b†]	95% CI; Diff. test: 1, 1[b]	Code, Matrix	No
Bos et al. (2018)	178	Clinical	40 (18–65)	55%	BDI-II	DSM-IV	20	2	Strength, Pred.	↓MDD; ↑GS; 3, 5 clusters	–	210	Polychoric	GGM	0.13, 0.13[b]	–	No	Gold
Boschloo et al. (2016)	422, 79[c]	Community	41.9 (18–65)	66.4%	IDS-C	DSM-IV	12	1	Strength	–	–	78	Polychoric	GGM	–	–	No	Gold
Cramer et al. (2016)	2096	Community	35.1 (18–60)	49.9%	Interview	DSM-III-R	14	4	Strength	–	–	105	Tetrachoric	Assoc.	–	–	Matrix	Green
de la Torre-Luque et al. (2020)	232, 195[d]	Community	75.8 (65–80+)	75%	CIDI	DSM-IV	11	2	Strength, Bet., Clo.	Ø GS	Different: Spain/Nigeria	66	Tetrachoric	MGM	–	95% CI	Matrix	No
Fried et al. (2016)	3463	Clinical	41 (18–75)	63%	IDS-C	DSM-IV	15, 28[e]	2	Strength, Bet., Clo.	–	–	120, 406[e]	Polychoric	GGM	–	–	No	No
Hakulinen et al. (2020)	5998, 595[c]	Community	(18–81+)	52%	BDI-II	DSM-IV	13	2	Strength, E.I., P.C., Pred.	Ø GS; 3, 4 clusters[c]	Similar: w/o depressive disorder	91	Polychoric	FGL	0.75, 0.67[c]	95% CI; Diff. test: 18, 5[c]	Code, Matrix	Gold
Hartung et al. (2019)	4020[f]	Community	56.5	51%	PHQ-9	DSM-IV	9	2	Strength, Pred.	↑MDD ↓GS	Different: cancer pts./ population	45	Spearman	uGGM	–	–	Code, Matrix	Green
Kendler et al. (2018)	5952	Community	(30-60)	100%	CIDI	DSM-IV	19	1	Strength	3 clusters	–	190	Tetrachoric	Ising	0.67	95% CI; Diff. test: 8	Code, Matrix	Green
Madhoo and Levine (2016)	2862, 2585, 2578[g]	Clinical	40.8 (18–75)	63.7%	IDS-SR	DSM-IV	14	3	Strength, Bet., Clo.	↓MDD; ↑GS	–	105	Polychoric	GGM	–	–	No	No
McElroy et al. (2019)	566, 227, 174[h]	Clinical	14.43 (8–18)	78%	RCADS	DSM-IV	10	6	Strength, Bet.	↓MDD; ↑GS[¥]	Similar: w/o response to treatment	55	Polychoric	GGM	–	Diff. test: 3,4,9,10,1,0[h]	No	Gold
McWilliams et al. (2017)	216	Community	47.06	62%	PHQ-9	DSM-IV	9	1	Strength, Bet., Clo.	–	–	45	Polychoric	GGM	–	–	No	No
Mullarkey et al. (2019)	1409	Community	14.35 (13–19)	52.8%	CDI	DSM-5	26	3	Strength, Bet., Clo.	Ø GS	Different: male/ female	351	Spearman	Ising	0.67	95% CI; Diff. test (7)	No	Green
Murri et al. (2018)	8557	Community	74	59.4%	EURO-D	DSM-5	12	5	Strength, Bet., Clo.	–	–	80	Polychoric	GGM	0.85	95% CI; Diff. test: 17	Matrix	No
Osborn et al. (2020)	2192	Community	15.21 (13–18)	56.8%	PHQ-9	DSM-5	8	3	Strength	↓MDD; ↑GS	Similar: low/ elevated symptoms	36	Polychoric	GGM	0.92	95% CI	No	Green
Park & Kim, 2020	223	Clinical	45.74 (18–65)	63.2%	BDI-II	DSM-5	21	1	Strength, Bet., Clo.	2 clusters	–	231	Polychoric	GGM	0.59	95% CI	No	No
Santos Jr et al. (2017)	515	Community	24.6 (18–40)	100%	CES-D	DSM-IV	20	1	Strength, Bet., Clo. Pred.	↑MDD; ↑GS	Similar: w/o MDD	210	Polychoric	GGM	0.28	95% CI; Diff. test: 0	Code, Matrix	Green
Schweren et al. (2018)	233, 232[i]	Clinical	(11–17)	75.1%	MFQ	DSM-5	11	2	Strength	↑MDD; ↑GS	–	66	Spearman	GGM	–	–	No	Green
		Clinical		65.1%	IDS-SR	DSM-IV	11	2		↑MDD; ↑GS	–	66	Spearman	GGM	–	–	No	Green

(*continued on next page*)

Table 1 (*continued*)

Reference	Sample characteristics				MDD symptoms			Network analysis									Reproducibility	
								Characteristics				Estimation			Stability			
	N	Study design	Age M (range)	Gender % female	Measure	Diagnostic manual	N	Model N[1]	Centrality & control	Symptoms connectivity[2]	Structure differences[3]	Parameters N	Correlation technique	Analytic model	CS coefficient	Edge accuracy[4]	Open Data	Open Access
van Borkulo et al. (2015)	262, 253[j]		40.9 (18–65)						Strength, Bet., Clo. E. C.									
Van Loo et al. (2018)	5784	Clinical	44.4 (30–60)	100%	CIDI	DSM-IV	24	1	–	Ø GS	Similar: w/o genetic or environment risk factors	287	Spearman	Ising	–	95% CI	Code	Green
Wasil, Venturo-Conerly, Shinde, Patel, & Jones, 2020	13,035	Community	13.8 (13–14)	47.5%	PHQ-9	DSM-5	9	1	E.I.	–	–	45	Polychoric	GGM	0.85[†]	95% CI	Code	Green

Note. [1]: Cross-sectional networks that met inclusion criteria for the review. [2]: ↑ MDD ↑ GS = MDD symptoms increased as Global Strength increased; ↓ MDD ↑ GS = MDD symptoms decreased as Global Strength increased; ↑ MDD ↓GS = MDD symptoms increased as Global Strength decreased; Ø GS= Not significant and no difference in Global Strength. [3]: Similar = no differences in network structures across different samples; Different = significant network structures differences across samples. [4]: 95% confidence intervals for edge-weight accuracy + Diff. Test: n = Bootstrapped Edges Comparison, with number of edges reliably different than two thirds of other edges in that network. [a]: Only in Stroke condition. [b]: Baseline vs. week 8 of SSRI treatment. [c]: MDD (-) vs. MDD (+). [d]: Spanish vs. Nigerian Sample. [e]: DSM vs. DSM and Non-DSM symptoms. [f]: General vs. cancer population. [g]: Baseline, endpoint, change. [h]: Improved (T1, T3), Unchanged (T1, T3), Deteriorated (T1, T3); [i] = Good vs. Poor SSRI Responders; [j] = Remittent vs. Persistent MDD. [†]: Substitution of EI CS-coefficient for Strength CS-coefficient. [¥]: Only significant for Improved and Unchanged Group (T1 and T3).

PHQ-9= Patient Health Questionnaire; IDS= Inventory of Depressive Symptomatology (SR: self-report; C: clinician); BDI-II= Beck Depression Inventory; CIDI= Composite International Diagnostic Interview CDI= Children's Depression Inventory; EURO-D= European Depression scale; MFQ= Mood and Feelings Questionnaire; MADRS: Montgomery Åsberg Depression Rating Scale; RCADS= Revised Children's Anxiety and Depression Scale. GGM= Gaussian Graphical Model; Ising= Ising Model; Assoc. = Association; uGGM= Unregularized Gaussian Graphical Model; FGL= Fused Graphical Lasso; MGM = Mixed Graphical Model. Bet. = Betweenness; Clo. = Closeness; E.I. = Expected Influence; P.C. = Participant Coefficient; E.C.= Eigenvector Centrality; Pred. = Predictability. Gold = Gold open access, Green = Green Open Access.

Park & Kim, 2020; Santos Jr et al., 2017; van Loo et al., 2018; Wasil, Venturo-Conerly, Shinde, Patel, & Jones, 2020) or two networks (n = 8; Airaksinen et al., 2020; Bos et al., 2018; de la Torre-Luque et al., 2020; Fried et al., 2016; Hakulinen et al., 2020; Hartung, Fried, Mehnert, Hinz, & Vehling, 2019; Schweren et al., 2018; van Borkulo et al., 2015). The rationale for constructing multiple networks consisted of making group-level comparisons (n = 10; An et al., 2019; Cramer, Borsboom, Aggen, & Kendler, 2012; de la Torre-Luque et al., 2020; Fried et al., 2016; Hakulinen et al., 2020; Hartung et al., 2019; Mullarkey et al., 2019; Murri et al., 2018; Osborn et al., 2020; Schweren et al., 2018), or measuring within-group MDD differences over time in the context of treatment interventions (n = 6; Airaksinen et al., 2020; Berlim, Richard-Devantoy, dos Santos, & Turecki, 2020; Bos et al., 2018; Madhoo & Levine, 2016; McElroy et al., 2019; van Borkulo et al., 2015).

Studies also examined global network strength (n = 15 studies; 42 networks), whose values were proportional to the number of network parameters and ranged from as low as 0.19 (8 Nodes - Osborn et al., 2020) to as high as 38.31 (26 Nodes - Mullarkey et al., 2019). The review sample also had varying findings related to the relationship between global strength and MDD. For instance, while some studies reported lower global strength to be associated with less symptom severity (e.g., An et al., 2019; Santos Jr et al., 2017; Schweren et al., 2018; van Borkulo et al., 2015), in others, higher global strength was linked to decreased symptom severity (e.g., Berlim et al., 2020; Bos et al., 2018; McElroy et al., 2019; Osborn et al., 2020). Global Strength characteristics for each study and their relevant study parameters are reported in Supplementary Table S2.

Community symptom structures within MDD were occasionally tested (n = 4 studies; 6 networks). Community detection results ranged from 2 (Park & Kim, 2020), 3 (Bos et al., 2018; Kendler et al., 2018; Hakulinen et al., 2020), 4 (Hakulinen et al., 2020), and 5 (Bos et al., 2018), with three communities as the most frequently reported. Moreover, four communities were demonstrated to be a significantly less stable than a three community solution, as based on sensitivity analyses by Hakulinen et al. (2020). Subsets denoted symptoms related to Cognitive Disturbances, Somatic Difficulties, and Mood: cognitive, affective, somatic (in Bos et al., 2018); neurovegetative/mood, cognitive, anxiety/irritability (in Kendler et al., 2018).

Groups were tested for differences in their MDD network structure (n = 10 studies; 31 networks), with significant differences primarily found in community samples. Specifically, between males and females in adolescent depression (Mullarkey et al., 2019), community-dwelling older adults from Spain and Nigeria (de la Torre-Luque et al., 2020), and cancer patients and the general population (Hartung et al., 2019). Only one clinical sample (Berlim et al., 2020) reported significant group differences: before and 8 weeks after an antidepressant treatment (escitalopram and desvenlafaxine). Conversely, no significant group differences were found between those with low and high anxiety (An et al., 2019), with and without MDD (Hakulinen et al., 2020; Santos Jr et al., 2017), low and elevated MDD symptoms (Osborn et al., 2020), response treatment (McElroy et al., 2019), and with and without genetic and environmental risk factors (van Loo et al., 2018).

3.1.4. Network estimation

Network parameters ranged from 36 to 406, with most studies (20 of the 23 total) having a subject-to-parameter ratio considered appropriate to accurately estimate stable network parameters (calculation is based on a rule of thumb; see: Fried & Cramer, 2017). Polychloric was the most popular network correlations method (n = 14 studies; n = 39 networks), while the most popular network estimation used was the GGM (n = 15 studies; 41 networks). All networks used the Fruchterman–Reingold algorithm as it's technique for node position and network visualization.

3.1.5. Network stability

In terms of the accuracy of node strength centralities, the most used stability measure was case dropping bootstrap (n = 13 studies; 37 networks), and its correlation stability (CS) coefficients (n = 8 studies; 18 networks). Of the studies that reported strength CS coefficients, the median value was 0.67, ranging from 0.13 to 0.92. Furthermore, two studies reported a CS coefficient for expected influence (Berlim et al., 2020; Wasil, Venturo-Conerly, Shinde, Patel, & Jones, 2020).

In terms of edge accuracy, bootstrapped confidence intervals for estimated edge weights were available for the majority of networks (n = 12 studies; 31 networks). Seven studies (24 networks) also used bootstrapped difference testing to estimate if their edges were significantly different.

3.1.6. Reproducibility

Within the review sample, 8 studies had open data in the form of reported data matrices, seven studies made available their scripts, and 16 studies were considered open access, of which 10 being Green open access and 6 Gold open access.

3.2. Node and edges comparisons

3.2.1. Symptoms strength centrality rankings

Strength was the most popular centrality index used, with 20 studies containing Strength node centrality. Of the resulting 50 cross-sectional networks, 17 were excluded as derived from subsamples already included in a broader network. In addition, we excluded one network (Madhoo & Levine, 2016) derived from within-group estimated change scores. The final sample consisted of 19 studies that produced 32 cross-sectional networks, which were compared based on Strength centrality scores for the nine symptoms of MDD. Some of the 19 studies contained more than one network: Airaksinen et al. (2020), before (a) and after (b) a stroke; Bos et al. (2018), before (a) and after (b) an 8-week SSRI treatment; de la Torre-Luque et al. (2020), Spanish (a) and Nigerian (b) samples; Hakulinen et al. (2020), with MDD (a) and without (b) MDD diagnosis; Hartung et al. (2019), cancer patients (a) and normal population (b); Madhoo and Levine (2016), treatment baseline (a) and endpoint (b); McElroy et al. (2019), time 1 and time 3 of treatment for improved (a + b), unchanged (c, d), and deteriorated (e, f) groups; Schweren et al. (2018), good (a) and poor (b) SSRI responders; van Borkulo et al. (2015), remittent (a) and persistent MDD (b). The majority of examined networks (19/32) omitted one or more MDD symptoms, particularly Suicidal thoughts and behaviors (13/32). From networks containing multiple items assessing the same DSM-5 MDD symptom (e.g., sleep-onset and mid-nocturnal insomnia), we included only the highest ranking element per symptom. Additionally, we included auxiliary items from four studies by matching them to MDD symptoms. Specifically, we considered Pessimism (in: An et al., 2019, Hakulinen et al., 2020, Murri et al., 2018), and Helpless (in: Kendler et al., 2018) as signs of Depressed Mood; and Unreactive mood as a sign of Anhedonia (in Kendler et al., 2018). Original wording, ranking, and correspondence to MDD for each item for each study are reported in Supplementary Table S3.

Fig. 2 displays the ranked node strength centrality of 32 networks; each depressive symptom is ranked in decreasing order of centrality, from 1 (highest centrality) to 9 (lowest possible centrality). Aggregate findings across studies suggest that Fatigue was most commonly the MDD symptom with the highest Strength centrality in 10/32 networks (median rank = 2; Interquartile range IQR: 1–3), closely followed by Mood as the most central symptom in 9/32 networks (median rank = 2; IQR: 1–4). Fatigue was also reported as one of the top three central symptoms more frequently than Depressed Mood (26 vs. 20 networks, respectively). Further, results indicated that Worthlessness (median rank = 3; IQR: 2–5.75) and Anhedonia (median rank = 3; IQR: 2–5) were often tertiary in highest centrality. Moreover, Suicidal Thoughts or Behaviors (median rank = 7; IQR: 6–9) appeared as one of the three lowest centralities while being reported in a total of 19 networks. In terms of lowest ranking, Weight changes (median rank = 7.5; IQR: 6–8) were consistently reported as the symptom with the lowest centrality within

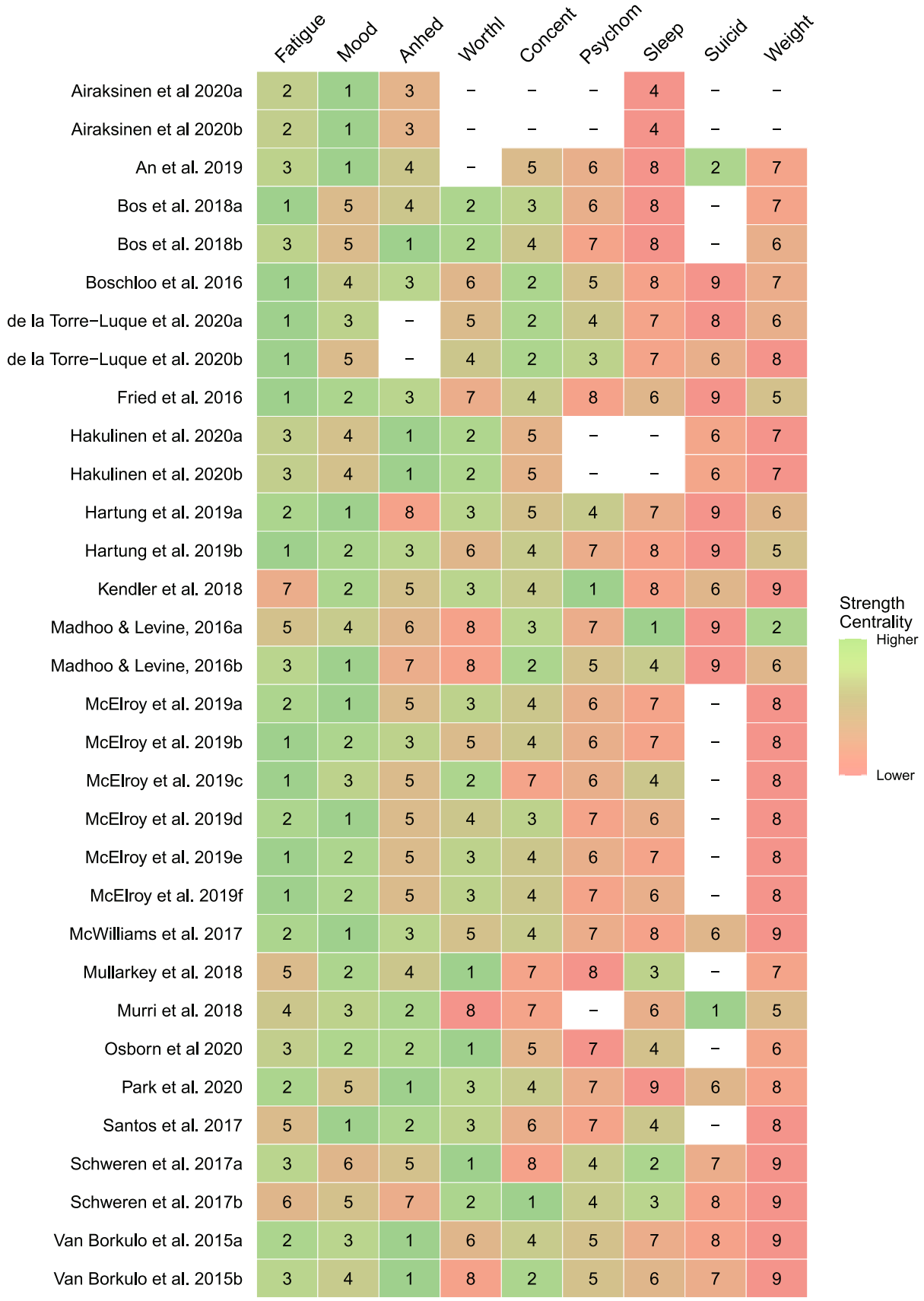

Fig. 2. Strength Centrality Rankings for cross-sectional networks of MDD symptoms (N=32). Numbers indicate Strength centrality rankings.

sixteen networks. Fig. 3 reports all median symptom rankings and their spread across the examined studies.

Lastly, Logistic regression analyses were performed to determine the role of sample size, type of sample (clinical vs. community), and if the sample was repeated (e.g., pre and post) as potential confounders of Fatigue and Mood being most central symptoms. For each network, rankings for Depressed Mood and Fatigue centrality were dichotomized based on if the symptom had the highest ranking Strength centrality of the other symptoms. Results from the logistic analyses indicated that no significant differences emerged in terms of sample characteristics for both Mood ($\chi 2$ (28, 31) = 3.29, p = .35; ROC AUC = 0.66; McFadden R^2 = 0.09) and Fatigue ($\chi 2$ (28, 31) = 0.94, p = .82; ROC AUC =0.60; McFadden R^2 = 0.02). Odds ratios and full estimates for each of the examined sample characteristics are reported in Supplementary Table S4.

3.2.2. Robust edges of MDD symptoms

Of the 23 examined studies, only seven reported edge weight difference testing, of which two were not included due to missing figure labeling (Mullarkey et al., 2019; Santos Jr et al., 2017). The final analytic sample consisted of twelve networks from five studies (Berlim et al., 2020; Hakulinen et al., 2020; Kendler et al., 2018; McElroy et al., 2019; Murri et al., 2018). We created a summary network graph based on the proportion of MDD robust edges in the twelve networks. An edge was included if positive, unique, and if it emerged as reliably different from at least two-thirds of the total edge weights within the same estimated network. Nodes were matched to MDD symptoms as per centrality rankings, with three studies containing either one (in: Kendler et al., 2018; McElroy et al., 2019) or two missing symptoms (in: Hakulinen et al., 2020). Only one edge was not included for being negative (i.e., describing deactivation patterns): Worthlessness-Fatigue (in: Hakulinen et al., 2020). The overall number of robust symptom edges per network is reported in Table 1, while the specific connections for each network and study are reported in Supplementary Table S5.

Fig. 4 illustrates the most frequently observed robust edges within MDD symptoms networks. The connection between Depressed Mood and Anhedonia was the most frequent robust edge (8/11 networks). Depressed Mood was also frequently connected to Worthlessness (5/12 networks), and the two symptoms shared a similar incidence of connections with Suicidal Thoughts and Behaviors (respectively 3/6 and 4/6 networks). Worthlessness-Suicidality was, along with edges associated with Mood (−Anhedonia, -Worthlessness, and – Suicidality), one of the

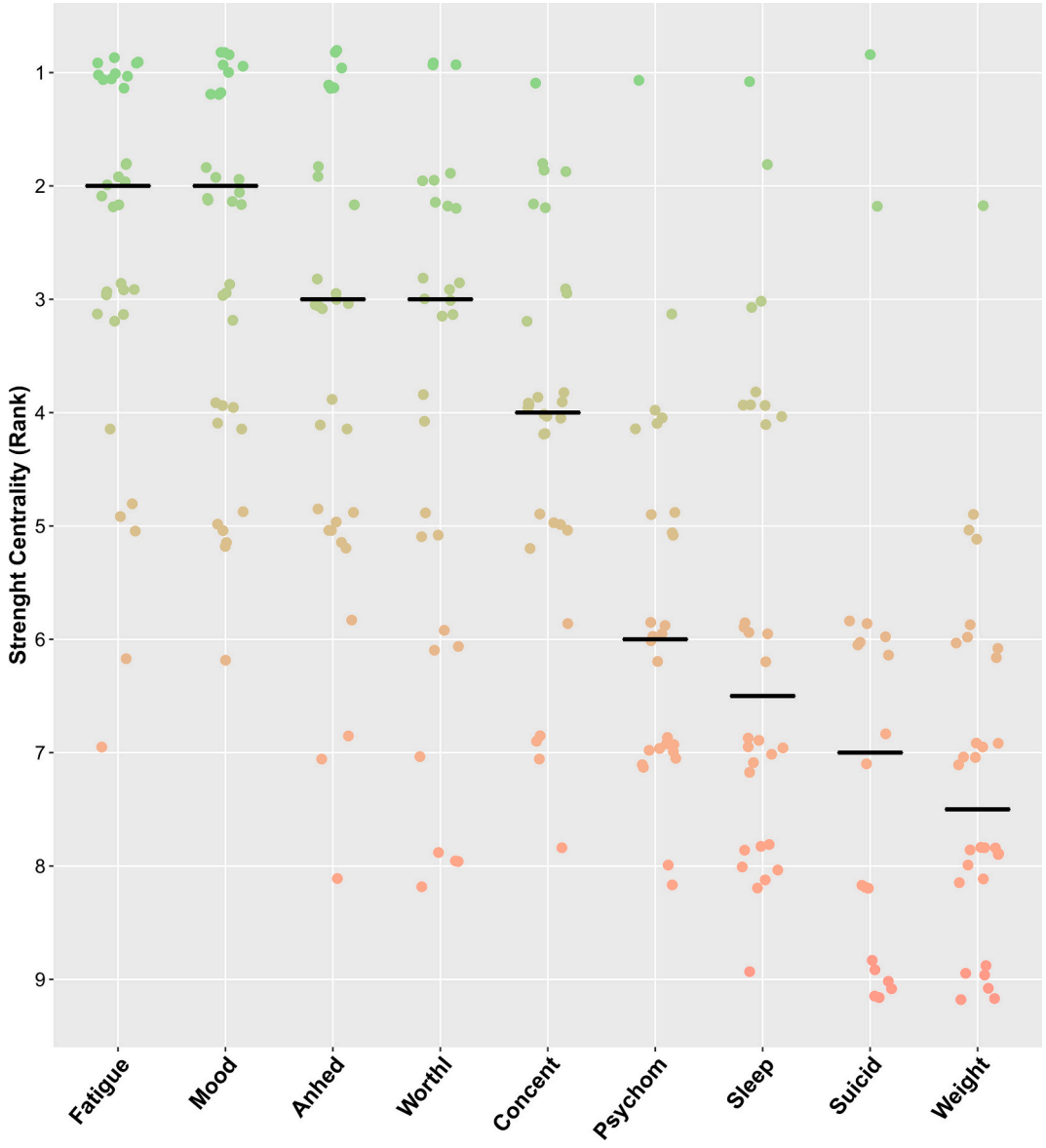

Fig. 3. MDD symptoms centrality median rankings for cross-sectional networks of MDD symptoms.

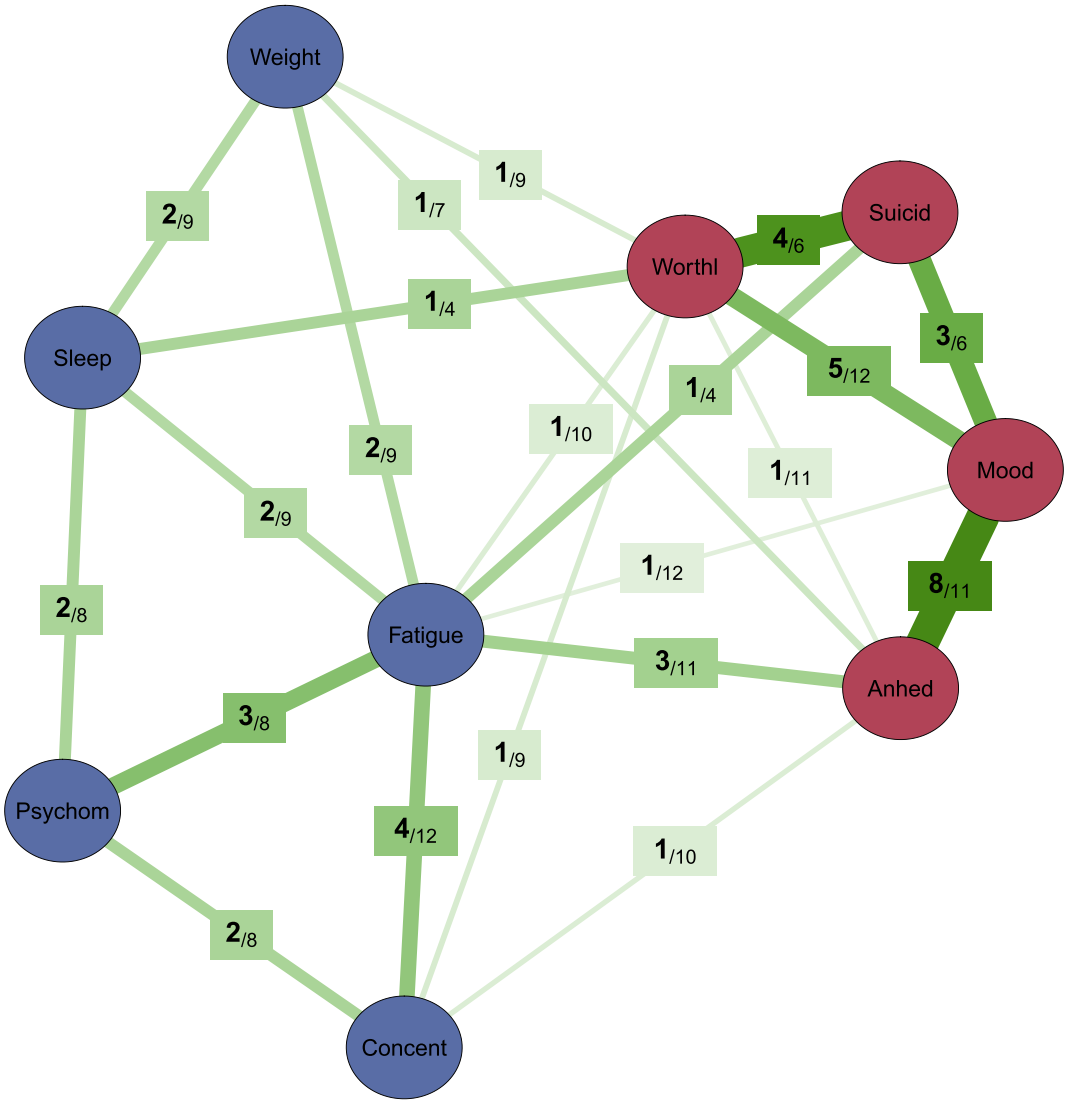

Fig. 4. Summary network of robust MDD edges. Edge-weights indicate proportion of robust symptoms connections across studies that included bootstrapped difference tests between edges.

only four edges found to be robust in more than 40% of the studies. Fatigue had the most direct robust connections with other MDD symptoms, particularly with Concentration difficulties (4/12 networks) and Psychomotor Disturbances (3/8 networks). Weight, Sleep and Fatigue were observed to have similar frequencies of connections (2/9 networks). The most robust connection between the affective and somatic communities was between Fatigue and Anhedonia (3/11 networks).

The spinglass algorithm clustered the robust edges network in two communities (906/1000 runs) of MDD symptoms, each characterized by more frequent robust connections among its nodes. Specifically, one cluster contained affective symptoms: Depressed Mood, Anhedonia, Worthlessness, and Suicidality; the other consisted primarily of somatic symptoms: Fatigue, Sleep, Weight, Psychomotor, as well as Concentration. When analyzing robust edges within the two communities, the highest number of robust connections were associated with Depressed Mood and Fatigue respectively. The most robust connection between the affective and somatic clusters was the edge between Anhedonia and Fatigue.

4. Discussion

Although depression is among the most commonly observed psychiatric disorders, its diagnosis and measurement have shown a worrisome level of heterogeneity. Network analysis has emerged as a highly promising approach to probe that heterogeneity and to better reveal the disorder's etiological underpinnings (Borsboom & Cramer, 2013). In this article, we report the first comprehensive review of network analyses for MDD. Three key findings emerged: 1) There is marked inconsistency in the assessments used to construct networks of MDD; 2) Within the constraint of measurement variability, Fatigue and Depressed Mood emerged as the most central symptoms; 3) Fatigue and Depressed Mood formed separated communities of robustly connected symptoms.

Our first key findings indicated that cross-sectional networks of MDD are constituted from a wide range of measurement scales, and as such, no gold standard to derive networks emerged. A potential cause could be that over 280 questionnaires have been used to measure depressive symptoms (Santor et al., 2006). This problem is exacerbated further when the choice of questionnaires used in different studies is determined by its slant toward a specific symptom subset. For instance, the Hamilton Rating Scale for Depression (HRSD) is often used for anti-depressant clinical trials because of its focus on somatic symptoms (Fried, 2017; Hamilton, 1960). Other scales may be chosen due to accounting for suitable non-DSM symptoms relevant to the study; for example, the

inclusion of diurnal mood variation and gastrointestinal problems within the IDS (Rush et al., 1986), or the incorporation of loneliness and disobedience within the CDI as symptoms of adolescent depression (Kovacs, 1992). In addition to the variance observed in measurement scales, we also observed significant variance within diagnostic frameworks. This raises whether specific discrepancies in network construction may be due in part to confounding inconsistencies in their diagnostic nosology (Borsboom et al., 2018).

Our second key finding related to the centrality hypothesis. Within the observed constraints of the measurement variability, as seen in Fig. 2 Fatigue and Depressed Mood were the most consistently central symptoms in MDD across studies. These results confirm previous research showing Fatigue as highly central (Cramer, Waldorp, Van Der Maas, & Borsboom, 2010; Fried et al., 2017) and as a potential target in informing of the onset of MDD (Contreras et al., 2019). Our findings also provide additional evidence that clinical interventions should consider monitoring or even targeting Fatigue, given that it could increase the prospect of reducing MDD symptoms activation. The fact that Depressed Mood emerged as a common central symptom is consistent with its status as a 'hallmark symptom' of MDD. The results of Anhedonia being listed as a tertiary centrality is more surprising, as Anhedonia often outperforms other depression symptoms in predicting depression (McMakin et al., 2012). Interestingly, the importance of Fatigue and Depressed Mood as emerged from their centrality rankings is consistent with their pivotal role in ICD-10's MDD diagnostic algorithm (World Health Organization, 1993), where at least one of the two symptoms is required in order to diagnose a depressive episode (i.e., F32.x criterion B). Conversely, we found Weight to have the lowest centrality across studies, which may have resulted from the symptom's lack of specificity (i.e., the weight symptom may indicate an increase or a decrease in appetite or weight). In sum, our review suggests that Fatigue and Depressed Mood are the two symptoms that most commonly influence interactions between other symptoms.

The clinical importance of highly central symptoms remains a hotly debated topic (Bringmann et al., 2019), given that the level of statistical associations between symptoms is influenced by multiple factors beyond clinical significance, as evinced from our findings on Suicidality. The symptom has a low base rate, is under-measured and under-reported (Nock et al., 2010), and therefore obtained low centrality scores - in steep contrast with its clinical relevance. Preliminary investigations on the prognostic utility of the most central depressive symptoms are, however, promising (Elliott, Jones, & Schmidt, 2020). While further longitudinal and time-intensive studies are warranted, the cross-sectional evidence here aggregated suggests that Fatigue and Depressed Mood, could serve as crucial mechanisms in the dynamics of MDD onset, maintenance, and, possibly, treatment.

Our third key finding revealed two communities of robust symptoms edges: (1) Depressed Mood, Worthlessness, Anhedonia, and Suicidality; (2) Fatigue, Concentration, and Psychomotor disturbances. Notably, each community included one of the two most central symptoms that emerged across studies (i.e., Fatigue and Depressed Mood). From a network psychometrics perspective, the two symptoms could dictate the organization of their MDD symptoms clusters. The two identified symptom clusters showed some similarity to previously identified MDD symptoms network communities (Bringmann, Lemmens, Huibers, Borsboom, & Tuerlinckx, 2015): Mood, Cognitive difficulties, and Somatic disturbances, respectively. Some distinction between the affective and somatic components of MDD is also alluded in the ICD-10 diagnostic guidelines, where the presence of Sleep, Weight, Psychomotor changes (but also Anhedonia) qualifies as somatic syndrome subtype (WHO, 1993). Moreover, differences in somatic symptoms were observed to be distinguishing two subtypes of depression (Lamers et al., 2013), a contrast further characterized by metabolic and inflammation differences. The link to inflammation could potentially reveal the role of Fatigue in depression networks, as inflammation causes both fatigue and lethargy. Interestingly, only 47% of depressed individuals show heightened inflammation (Kiecolt-Glaser, Derry, & Fagundes, 2015), offering further suggestion that the importance of Fatigue and Mood in their communities could represent distinct activation pathways. Combining findings linking the onset of inflammation with the development of depression with our key findings suggests that Fatigue and Mood, based on their high centrality and robust connections, could exert different pathways with a high degree of influence within their respective MDD symptoms sub-set. Also, despite their high centrality, Fatigue and Depressed Mood showed little robust connections to each other, possibly underscoring their unique role in the configuration of MDD symptoms.

In terms of specific robust edges, the connection between Mood and Anhedonia was the most frequent edge across the examined networks. The association between the two symptoms is well-documented and suggested to be related to the dopaminergic pathway and its role in triggering MDD (Stein, 2008). Worthlessness showed robust connections with other affective symptoms, a prominence consistent with the high comorbid rates of MDD with conditions involving experiences of guilt, particularly post-traumatic (Lee, Scragg, & Turner, 2001), obsessive-compulsive (Shafran, Watkins, & Charman, 1996), and alcohol/substance use disorders (Lux & Kendler, 2010; Treeby & Bruno, 2012). Outside of edges in the affective symptoms cluster, the connections between the other symptoms were robust only in less than 40% of the examined networks. While the lack of consistent robust edges echoes debates on the replicability of psychological networks between studies (e.g., Forbes et al., 2017), we believe it to be engendered by study factors, rather than by the limitations of network psychometrics. First, by examining robust edges (significantly different than 2/3 of other edges) we prioritized specificity over sensitivity, to avoid including spurious connections. Secondly, different studies used different measures for MDD, in particular splitting somatic symptoms in different nodes. The low robustness of somatic symptom edges could, therefore, once again be attributed to the heterogeneity of the examined studies, which measured different elements despite defining it under the same MDD construct (Fried & Nesse, 2015a, 2015b). Lastly, it is also worth noting that the connection between Anhedonia and Fatigue was the most robust edge bridging the affective and somatic symptoms clusters. The co-occurrence of Fatigue and Anhedonia is well known, as both exhaustion and lack of participation in activities lead to behavioral deactivation, which reinforces depressive states (Billones, Kumar, & Saligan, 2020).

Our review boasts several strengths. It was the first to systematically evaluate the results of existing network analyses on MDD symptoms. As such, the review was also the first to construct summary figures of centrality and edge weights, which made it possible to objectively determine the most central symptoms and edge connections within MDD and across studies. Thus, our review's summary findings offer valuable recommendations for future studies relating to methodology, research directions, and replicability. Concerning methodology, network studies should include multiple centrality indexes to supplement strength centrality, including expected influence and (if testing for discrete communities within a network) participant co-efficient indices, as well as bridge statistics (Jones, Ma, & McNally, 2019) to study associations with other conditions throughout the MDD network. Concerning research directions, more effort is needed to reduce between-study heterogeneity. A possible solution would be to limit the number of clinical measures and diagnostic frameworks used to assess MDD. This approach is consistent with recent NIMH efforts that require the use of the PHQ-9 as primary depression measure (Patalay & Fried, 2020), thus eliminating auxiliary elements and the need of an item conversion process to fit DSM-5 criteria. It is, however, a research decision not without compromise, as depression scales measure different symptoms which often do not overlap (Fried & Nesse, 2015a, 2015b). A temporary methodological solution, while building consensus on the most appropriate diagnostic framework, would be to limit the presence of auxiliary nodes analytically using the goldbricker procedure. This tool (for

technical details see:(Jones, 2018)) is used to test statistical differences between edges within a network, and could indicate if multiple nodes may be redundant as representing the same underlying MDD construct. Concerning replicability, stability measures should be mandatory to improve the generalizability of network model findings. It is strongly encouraged for any study to contain bootstrapped confidence intervals of all edge weights, centrality stability indexes, and correlation stability coefficients. Lastly, researchers network investigations should attempt to include as many reproducibility signs as possible, hence following an Open Science Framework (Foster & Deardorff, 2017).

It is important to note, however, that the potential contribution and implications of our review must be couched in the context of several limitations. There are currently no established methodologies for aggregating and comparing findings from different network studies. As such, although bias was limited by strict review protocol, data collection and the construction of summary figures were done by tallying centrality and edge weights reported for a specific network. Comparison of symptom centrality was limited by the sole availability across the examined literature of ranking data, which describes the order of elements but not their relative distance (possibly consisting of minor differences). For this reason, we focused our discussion only to the most central symptoms as trended across the 32 examined cross-sectional networks. In terms of comparing connections between symptoms, we examined the presence of edges between two symptoms and not their weights. While we limited our evaluation to studies where edge stability metrics were available, and all these studies had sufficient subjects to ensure parameter stability (Fried & Cramer, 2017), the presence/absence of edges could have been influenced by multiple study factors that are not accounted by edge counts. As a measure of additional rigor, we focused on robust edges (significantly different than 2/3 of the remaining edges in bootstrap comparison tests), which was suggested by a previous network review (Birkeland et al., 2020). Nevertheless, future studies should further assess our findings on MDD networks by incorporating testing for effect sizes of network models and their corresponding results. The construction of meta-analyses for networks is anticipated mainly through the forthcoming Meta-analytic Gaussian Network Aggregation (MAGNA) framework (Epskamp, Isvoranu, & Cheung, 2020), which is currently under review. Importantly, findings of a previous examination on centrality rankings and robust edges of PTSD (Birkeland et al., 2020) held to meta-analytic scrutiny (Isvoranu, Epskamp, & Cheung, 2020) and found similar results in terms of robust edges and symptoms relative importance. Nevertheless, as a meta-analytic approach, MAGNA would offer a higher level of evidence. To aggregate network findings across MDD studies using MAGNA, covariance matrices of each network would be needed: MAGNA meta-analysis does not aggregate results but rather estimates a new combined GGM using structural equation modeling (Epskamp et al., 2020). As noted in our systematic review, only a handful of studies made their covariance matrices available, which hinders meta-analytic efforts based on currently available data. As network covariance matrices are obtainable when estimating networks through packages such as qgraph (Epskamp et al., 2012), future studies should incorporate them in their supplementary materials whenever possible. This would help future efforts to consolidate our understanding of cross-sectional networks of MDD, as well as of other conditions.

There are also a number of other limitations in our study. Firstly, we limited our review to studies that contained cross-sectional data and excluded any temporal or longitudinal data. Therefore, causality cannot be directly inferred, limiting the ability to make generalized inferences on the dynamics of MDD and individuals. Secondly, by summarizing network data, a layer of specificity was diminished. In other words, by clustering symptoms into distinct categories related to the DSM, the possibility to understand the temporal dynamics of symptom appearance is lost. With this approach, we could not, for instance, distinguish between early, middle, or late insomnia or, as is required by diagnostic category, separate insomnia and hypersomnia. This limitation is particularly noteworthy when considering psychomotor disturbance, given that psychomotor retardation is significantly more detrimental than psychomotor agitation (Fried & Nesse, 2014).

In conclusion, this review consolidated network studies to determine findings, distinguish gaps within the literature, and propose future directions for the field of depression research. Despite the need for temporal data to make causal inferences, cross-sectional networks may still be crucial in discovering the initial interactions that elucidate how specific symptoms and connections may significantly alter the course of MDD. The field of depression research still has much to uncover, and network psychometrics are a promising methodology that could demystify its ontology and perhaps inform its prevention. .

Contributors

MM, AC, GB jointly designed the manuscript's Conceptualization, Methodology, Writing - original draft, and Writing - review & editing. AC oversaw Data curation. MM conducted Visualization and Analysis. All authors have approved the final manuscript.

Funding

Matteo Malgaroli is supported by the Scholars Program award from NYU Langone Clinical and Translational Science Institute, and by the American Foundation for Suicide Prevention research grant PRG-0-104-19. George A. Bonanno is the recipient of NIH grant R01MH091034. Funding agencies had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

Declaration of Competing Interest

No potential conflict of interest was reported by the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2021.102000.

References

Airaksinen, J., Gluschkoff, K., Kivimäki, M., & Jokela, M. (2020). Connectivity of depression symptoms before and after diagnosis of a chronic disease: A network analysis in the US health and retirement study. *Journal of Affective Disorders, 266*, 230–234.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

An, M. H., Park, S. S., You, S. C., Park, R. W., Kim, H. K., & Son, S. J. (2019). Depressive symptom network associated with comorbid anxiety in late-life depression. *Frontiers in Psychiatry, 10*, 856.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio, 78*(2), 490–498.

Berlim, M. T., Richard-Devantoy, S., dos Santos, N. R., & Turecki, G. (2020). The network structure of core depressive symptom-domains in major depressive disorder following antidepressant treatment: a randomized clinical trial. *Psychological Medicine*, 1–15.

Billones, R. R., Kumar, S., & Saligan, L. N. (2020). Disentangling fatigue from anhedonia: A scoping review. *Translational Psychiatry, 10*(1), 1–11.

Birkeland, M. S., Greene, T., & Spiller, T. R. (2020). The network approach to posttraumatic stress disorder: A systematic review. *European Journal of Psychotraumatology, 11*(1), 1700614.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks, 29*(4), 555–564.

van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry, 72*(12), 1219–1226.

van Borkulo, C. D., Boschloo, L., Kossakowski, J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2017). *Comparing network structures on three aspects: A permutation test* (Manuscript submitted for publication).

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry, 16*(1), 5–13.

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9*, 91–121.

Borsboom, D., Rhemtulla, M., Cramer, A. O. J., van der Maas, H. L., Scheffer, M., & Dolan, C. V. (2016). Kinds versus continua: A review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological Medicine, 46*(8), 1567–1579.

Borsboom, D., Robinaugh, D. J., Group, T. P, Rhemtulla, M., & Cramer, A. O. (2018). Robustness and replicability of psychopathology networks. *World Psychiatry, 17*(2), 143.

Bos, F. M., Fried, E. I., Hollon, S. D., Bringmann, L. F., Dimidjian, S., DeRubeis, R. J., & Bockting, C. L. (2018). Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. *Social Psychiatry and Psychiatric Epidemiology, 53*(6), 617–627.

Boschloo, L., van Borkulo, C. D., Borsboom, D., & Schoevers, R. A. (2016). A prospective study on how symptoms in a network predict the onset of depression. *Psychotherapy and Psychosomatics, 85*(3), 183–184.

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., … Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology, 128*(8), 892.

Bringmann, L. F., Lemmens, L. H. J. M., Huibers, M. J. H., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the Beck depression inventory-II. *Psychological Medicine, 45*(4), 747–757.

Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics, 88*(2), 71–83.

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2019). Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences, 136*, 68–78.

Cramer, A. O., Borsboom, D., Aggen, S. H., & Kendler, K. S. (2012). The pathoplasticity of dysphoric episodes: Differential impact of stressful life events on the pattern of depressive symptom inter-correlations. *Psychological Medicine, 42*(5), 957–965.

Cramer, A. O., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLoS One, 11*(12), Article e0167490.

Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences, 33*(2–3), 137–150.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695*(5), 1–9.

Elliott, H., Jones, P. J., & Schmidt, U. (2020). Central symptoms predict posttreatment outcomes and clinical impairment in anorexia nervosa: A network analysis. *Clinical Psychological Science, 8*(1), 139–154.

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(4), 1–18.

Epskamp, S., Isvoranu, A., & Cheung, M. W. (2020, April 1). *Meta-analytic Gaussian network aggregation*. https://doi.org/10.31234/osf.io/236w8.

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLoS One, 12*(6).

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research, 53*(4), 453–480.

Forbes, M. K., Wright, A. G., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology, 126*(7), 969.

Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association, 105*(2), 203.

Fried, E. I. (2015). Problematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. *Frontiers in Psychology, 6*, 309.

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders, 208*, 191–197.

Fried, E. I., & Cramer, A. O. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science, 12*(6), 999–1020.

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are'good'depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders, 189*, 314–320.

Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLoS One, 9*(2).

Fried, E. I., & Nesse, R. M. (2015a). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR* D study. *Journal of Affective Disorders, 172*, 96–102.

Fried, E. I., & Nesse, R. M. (2015b). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine, 13*(1), 72.

Fried, E. I., van Borkulo, C. D., Cramer, A. O., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology, 52*(1), 1–10.

Hakulinen, C., Fried, E. I., Pulkki-Råback, L., Virtanen, M., Suvisaari, J., & Elovainio, M. (2020). Network structure of depression symptomology in participants with and without depressive disorder: The population-based health 2000–2011 study. *Social Psychiatry and Psychiatric Epidemiology*, 1–10.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*(1), 56.

Hartung, T. J., Fried, E. I., Mehnert, A., Hinz, A., & Vehling, S. (2019). Frequency and network analysis of depressive symptoms in patients with cancer compared to the general population. *Journal of Affective Disorders, 256*, 295–301.

Haslbeck, J., Ryan, O., Robinaugh, D., Waldorp, L., & Borsboom, D. (2019). *Modeling psychopathology: From data models to formal theories*.

Haslbeck, J., & Waldorp, L. J. (2020). mgm: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data. *Journal of Statistical Software, 93*(8), 1–46. https://doi.org/10.18637/jss.v093.i08.

Haslbeck, J. M., & Waldorp, L. J. (2018). How well do network models predict observations? On the importance of predictability in network models. *Behavior Research Methods, 50*(2), 853–861.

Isvoranu, A. M., Epskamp, S., & Cheung, M. (2020). Network models of post-traumatic stress disorder: A meta-analysis. *PsyArXiv Preprint*. https://doi.org/10.31234/osf.io/8k4u6.

Jones, P. J. (2018). *networktools: Tools for identifying important nodes in networks. R package version 1.2.0.*. Retrieved form https://CRAN.R-project.org/package=networktools.

Jones, P. J., Heeren, A., & McNally, R. J. (2017). Commentary: A network theory of mental disorders. *Frontiers in Psychology, 8*, 1305.

Jones, P. J., Ma, R., & McNally, R. J. (2019). Bridge centrality: A network approach to understanding comorbidity. *Multivariate Behavioral Research*, 1–15.

Kendler, K. S., Aggen, S. H., Flint, J., Borsboom, D., & Fried, E. I. (2018). The centrality of DSM and non-DSM depressive symptoms in Han Chinese women with major depression. *Journal of Affective Disorders, 227*, 739–744.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA, 289*(23), 3095–3105.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 617–627.

Kessler, R. C., & Üstün, T. B. (2004). The world mental health (WMH) survey initiative version of the world health organization (WHO) composite international diagnostic interview (CIDI). *International Journal of Methods in Psychiatric Research, 13*(2), 93–121.

Kiecolt-Glaser, J. K., Derry, H. M., & Fagundes, C. P. (2015). Inflammation: Depression fans the flames and feasts on the heat. *American Journal of Psychiatry, 172*(11), 1075–1091.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509–515.

Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E, 80*(5), Article 056117.

Lee, D. A., Scragg, P., & Turner, S. (2001). The role of shame and guilt in traumatic events: A clinical model of shame-based and guilt-based PTSD. *British Journal of Medical Psychology, 74*(4), 451–466.

Letina, S., Blanken, T. F., Deserno, M. K., & Borsboom, D. (2019). Expanding network analysis tools in psychological networks: Minimal spanning trees, participation coefficients, and motif analysis applied to a network of 26 psychological attributes. *Complexity, 2019*.

van Loo, H. M., Van Borkulo, C. D., Peterson, R. E., Fried, E. I., Aggen, S. H., Borsboom, D., & Kendler, K. S. (2018). Robust symptom networks in recurrent major depression across different levels of genetic and environmental risk. *Journal of Affective Disorders, 227*, 313–322.

Lux, V., & Kendler, K. S. (2010). Deconstructing major depression: A validation study of the DSM-IV symptomatic criteria. *Psychological Medicine, 40*(10), 1679–1690.

Madhoo, M., & Levine, S. Z. (2016). Network analysis of the quick inventory of depressive symptomatology: Reanalysis of the STAR* D clinical trial. *European Neuropsychopharmacology, 26*(11), 1768–1774.

McElroy, E., Napoleone, E., Wolpert, M., & Patalay, P. (2019). Structure and connectivity of depressive symptom networks corresponding to early treatment response. *EClinicalMedicine, 8*, 29–36.

McMakin, D. L., Olino, T. M., Porta, G., Dietz, L. J., Emslie, G., Clarke, G., … Shamseddeen, W. (2012). Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment–resistant depression. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(4), 404–411.

McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy, 86*, 95–104.

McWilliams, L. A., Sarty, G., Kowal, J., & Wilson, K. G. (2017). A network analysis of depressive symptoms in individuals seeking treatment for chronic pain. *The Clinical Journal of Pain, 33*(10), 899–904.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery, 8*(5), 336–341.

Mullarkey, M. C., Marchetti, I., & Beevers, C. G. (2019). Using network analysis to identify central symptoms of adolescent depression. *Journal of Clinical Child & Adolescent Psychology, 48*(4), 656–668.

Murri, M. B., Amore, M., Respino, M., & Alexopoulos, G. S. (2018). The symptom network structure of depressive symptoms in late-life: Results from a European population study. *Molecular Psychiatry*, 1–10.

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological Science, 21*(4), 511–517.

Osborn, T. L., Campbell, S., Ndetei, D., & Weisz, J. (2020). *Network analysis reveals central symptoms of adolescent depression and anxiety in sub-Saharan Africa* (Manuscript submitted for publication).

Park, S. C., & Kim, D. (2020). The Centrality of Depression and Anxiety Symptoms in Major Depressive Disorder Determined Using a Network Analysis. *Journal of Affective Disorders, 271*, 19–26. https://doi.org/10.1016/j.jad.2020.03.078.

Patalay, P., & Fried, E. I. (2020). Prescribing measures: Unintended negative consequences of mandating standardized mental health measurement. *The Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.13333. Epub ahead of print.

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry, 170*(1), 59–70.
Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E, 74*(1), Article 016110.
Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2019). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 1–14.
Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology, 125*(6), 747.
Rush, A. J., Giles, D. E., Schlesser, M. A., Fulton, C. L., Weissenburger, J., & Burns, C. (1986). The inventory for depressive symptomatology (IDS): Preliminary findings. *Psychiatry Research, 18*(1), 65–87.
Santor, D. A., Gregus, M., & Welch, A. (2006). FOCUS ARTICLE: Eight decades of measurement in depression. *Measurement: Interdisciplinary Research and Perspectives, 4*(3), 135–155.
Santos, H., Jr., Fried, E. I., Asafu-Adjei, J., & Ruiz, R. J. (2017). Network structure of perinatal depressive symptoms in Latinas: Relationship to stress and reproductive biomarkers. *Research in Nursing & Health, 40*(3), 218–228.
Schweren, L., van Borkulo, C. D., Fried, E., & Goodyer, I. M. (2018). Assessment of symptom network density as a prognostic marker of treatment response in adolescent depression. *JAMA Psychiatry, 75*(1), 98–100.
Shafran, R., Watkins, E., & Charman, T. (1996). Guilt in obsessive-compulsive disorder. *Journal of Anxiety Disorders, 10*(6), 509–516.
Stein, D. J. (2008). Depression, anhedonia, and psychomotor symptoms: The role of dopaminergic neurocircuitry. *CNS Spectrums, 13*(7), 561–565.
de la Torre-Luque, A., Ojagbemi, A., Caballero, F. F., Lara, E., Moreno-Agostino, D., Bello, T., … Ayuso-Mateos, J. L. (2020). Cross-cultural comparison of symptom networks in late-life major depressive disorder: Yoruba Africans and the Spanish population. *International Journal of Geriatric Psychiatry, 35*, 1060–1068.
Treeby, M., & Bruno, R. (2012). Shame and guilt-proneness: Divergent implications for problematic alcohol use and drinking to cope with anxiety and depression symptomatology. *Personality and Individual Differences, 53*(5), 613–617.
Wasil, A. R., Venturo-Conerly, K. E., Shinde, S., Patel, V., & Jones, P. J. (2020). Applying network analysis to understand depression and substance use in Indian adolescents. *Journal of Affective Disorders, 265*, 278–286.
Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.
World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research* (Vol. 2). World Health Organization.
World Health Organization. (2017). *Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2)*. World Health Organization.
Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry, 56*, 29–34.

Matteo Malgaroli is a Research Assistant Professor in the Department of Psychiatry at the NYU Grossman School of Medicine. He received his Ph.D. from Columbia University in 2018. His research focuses on machine learning methods to study psychiatric symptoms and their digital behavioral markers. He also studies digital health behavioral interventions (as an ideal source of large-scale clinical data), and teaches computational methods to psychology graduate students, psychiatry residents, and fellows.

Adam Calderon is a graduate student in the Loss Trauma and Emotion Lab at Columbia University, under the mentorship of Dr. George Bonanno. He received a B.S. in Behavioral Neuroscience from Quinnipiac University and an M.A. in Clinical Psychology from Columbia University. He is the recipient of a Fulbright research grant, a National Science Foundation REU scholar, and an APA graduate intern at the United Nations. His research focus includes the cognitive, affective, and psychophysiological mechanisms of emotion regulatory flexibility.

George A. Bonanno is a Professor of Clinical Psychology at Columbia University. He received his Ph.D. from Yale University in 1991. His research and scholarly interests have centered on the question of how human beings cope with loss, trauma and other forms of extreme adversity, with an emphasis on resilience and the salutary role of flexible coping and emotion regulatory processes. Professor Bonanno's recent empirical and theoretical work has focused on defining and documenting adult resilience in the face of loss or potential traumatic events, and on identifying the range of psychological and contextual variables that predict both psychopathological and resilient outcomes.