Structural Differences in REM and Non-REM Dream Reports Assessed by Non-Semantic Speech Graph Analysis

Joshua Martin

Student Registration: 20161014268

A dissertation submitted in partial fulfilment of the requirements for the award of the degree of Master of Science in Psychobiology

Department of Psychobiology
Federal University of Rio Grande do Norte (UFRN)

2017

Supervisor: Prof. Sidarta Ribeiro

Co-Supervisor: Prof. John Araújo
Martin, Joshua Michael.
Structural Differences in REM and Non-REM Dream Reports
Assessed by Non-Semantic Speech Graph Analysis / Joshua Michael
88 f.: il.

Dissertação (Mestrado) - Universidade Federal do Rio Grande
do Norte. Centro de Biociências. Programa de Pós-Graduação em
Psicobiologia.
Orientador: Prof. Dr. Sidarta Ribeiro.
Coorientador: Prof. Dr. John Fontenele Araújo.

1. Sonhos - Dissertação. 2. Sonhos não-REM - Dissertação. 3.
Estrutura do sonho - Dissertação. 4. Comprimento do relato -
Dissertação. 5. Análise de grafos - Dissertação. I. Ribeiro,
Sidarta. II. Araújo, John Fontenele. III. Universidade Federal
do Rio Grande do Norte. IV. Título.
**Título:** Structural Differences in REM and Non-REM Dream Reports Assessed by Non-Semantic Speech Graph Analysis

**Título em Português:** Diferenças Estrutrais nos Relatos de Sonho do Sono REM e Não-REM Avaliados por Análise de Grafos

**Autor:** Joshua Michael Martin

**Data da defesa:** 08/12/17

**Banca Examinadora:**

_____________________________________
Prof. Fernando Louzada
Universidade Federal do Paraná, UFPR, PR

_____________________________________
Prof. Adriano Tort
Universidade Federal do Rio Grande do Norte, RN

_____________________________________
Prof. Sidarta Ribeiro
Universidade Federal do Rio Grande do Norte,
Acknowledgements

I would like to thank all of the people who have helped make this work possible: to my supervisors, Sidarta Ribeiro and John Araújo, mentors and collaborators, Natalia Mota and Sergio Mota-Rolim, and external collaborators, Danyal Wainstein and Mark Solms, who have all played an integral role in my writing and ideas presented here. I would also like to thank CAPES for the scholarship and opportunity to complete a Master’s degree in Brazil. Finally, a special thanks to Marcelle Guil for being by my side and for the constant help and support over the past two years.
# Table of Contents

Abstract .............................................................................................................................................. 7

Resumo .................................................................................................................................................. 8

Introduction ......................................................................................................................................... 10

Differences in REM and Non-REM Dreaming ................................................................................. 12
- Recall rates ...................................................................................................................................... 14
- Report length ................................................................................................................................. 15
- Dream quality ............................................................................................................................... 16
- Structural organisation and narrative complexity ........................................................................ 17
- Summary ......................................................................................................................................... 18

Analysis of Speech Graph Attributes (SGA) .................................................................................. 19

Aims and Objectives .......................................................................................................................... 22

Methods ............................................................................................................................................ 24

Participants ......................................................................................................................................... 24
- Inclusion criteria ............................................................................................................................ 24
- Exclusion criteria ............................................................................................................................ 25

Procedure ........................................................................................................................................... 26
- Sleep study ...................................................................................................................................... 26
- Experimental night protocol ......................................................................................................... 26
- Awakening protocol ....................................................................................................................... 26
- Dream report collection .................................................................................................................. 27
Materials ................................................................................................................................. 28

Speech Graph Analysis (SGA) ................................................................................................. 28

Total Recall Count (TRC) .......................................................................................................... 32

Perception-Interaction Rating Scale (PIRS) ........................................................................... 33

Polysomnography (PSG) ........................................................................................................... 34

Voice recorder .......................................................................................................................... 34

Ethics and Informed Consent .................................................................................................. 34

Data Analyses ........................................................................................................................... 35

Wilcoxon sign-rank tests ........................................................................................................... 35

Generalised linear models ....................................................................................................... 35

Cumulative link models .......................................................................................................... 36

Significance values and goodness of fit .................................................................................... 37

Controlling for family-wise error rate .................................................................................... 37

Effect sizes ................................................................................................................................ 38

Random effects and independence of observations .................................................................. 38

Results .................................................................................................................................... 40

Descriptive Statistics ................................................................................................................ 40

Participants ............................................................................................................................... 40

Dream recall .............................................................................................................................. 40

Dream complexity ..................................................................................................................... 41

Inferential Statistics ................................................................................................................... 44

Differences in graph structure and TRC .................................................................................. 44
Testing for time of night effect .................................................................47
Distinguishing sleep stage based on graph structure and TRC .........................49
Testing the relationship to dream complexity ...............................................53
Dependence on dream complexity in predicting sleep stage .........................56
Model checking .......................................................................................56

Discussion ...............................................................................................58
REM Reports Are Longer and Have Larger Connectedness Compared to N2 ........58
Graph Connectedness in Relation to Dream Reports Across the Sleep Cycle ..........60
SGA as a Method for Dream Research ..........................................................64
Limitations and Perspectives for the Future ..................................................65

Conclusion ..............................................................................................68

References ...............................................................................................69

Supplementary Material .............................................................................77

Appendix A: An Illustration of the Sliding Window Method ..............................78
Appendix B: An Illustration of Randomised Word Graphs .................................79
Appendix C: Confirmation of Ethics Clearance ..............................................80
Appendix D: Supplementary Statistics ..........................................................81
Abstract

The extent to which Rapid Eye Movement Sleep (REM) mentation may differ to that of non-REM remains an important area of enquiry in dream research. Previous studies have found that dream reports collected after REM awakenings are, on average, longer, more vivid, bizarre, emotional and story-like compared to those collected after non-REM. Despite this, a comparison of the word-to-word structural organisation of dream reports is lacking, and traditional measures that distinguish REM and non-REM dreaming may be confounded by report length. The analysis of speech as directed word graphs can be suitably applied, as it provides a structural assessment of verbal reports, while controlling for differences in verbosity. In the present study, we aimed to investigate the differences in the connectedness of dream reports and their approximation to a random-like structure through applying speech graph analysis to 125 mentation reports obtained from 19 participants in controlled laboratory awakenings from REM and N2 sleep. We found that: (1) transformed graphs from REM possess a larger connectedness compared to those from N2; (2) measures of graph structure can predict ratings of dream complexity, where increases in connectedness and decreases in their random-like nature are observed in relation to increasing dream report complexity; and (3) the Largest Connected Component (LCC) can improve a model containing report length in predicting sleep stage and dream complexity. These results suggest that REM dream reports have a larger connectedness compared to N2 (i.e. words recur with a longer range), which we interpret to be related to underlying differences in dream complexity. They also point to speech graph analysis as a promising method for dream research, due to its relation to dream complexity and its potential to complement report length in dream analysis.

Key Words: dreams, non-REM dreaming, dream structure, report length, speech graph analysis.
**Resumo**

A diferença entre a mentação experimentada durante o sono de oculares rápidos (REM) e o sono não-REM persiste como questão importante para investigação no campo de pesquisa dos sonhos. Estudos anteriores têm mostrado que os relatos de sonho documentados depois do REM são, em média, mais longos, vividos, bizarros, emocionais e com aspectos mais narrativos do que os relatos do não-REM. Apesar desses achados, falta uma comparação estrutural entre relatos de sonho do REM e não-REM no que diz respeito à organização de palavra-a-palavra, e diversas medidas tradicionais de sonhos podem ser confundidas pelo comprimento do relato. A análise de fala transformada em grafos direcionados de palavras pode ser aplicada para fazer uma avaliação estrutural de relatos verbais e também para controlar as diferenças individuais de verbosidade.

No presente estudo, tivemos como objetivo investigar as possíveis diferenças na conectividade dos relatos e sua aproximação a uma estrutura aleatória através da análise de grafos em 125 relatos de sonho obtidos por 19 participantes em despertares controlados nas fases de sono REM e N2. Constatou-se que: (1) grafos do REM possuem uma conectividade maior do que os do N2; entretanto, essas diferenças não foram refletidas na aproximação a um grafo randômico; (2) diversas medidas de grafo podem predizer avaliações externas da complexidade do sonho, onde a conectividade aumenta e sua natureza randômica cai em relação à complexidade do relato; e (3) o Componente Maior Conectado (LCC) do grafo pode melhorar o ajuste de um modelo contendo o comprimento do relato como variável no discernimento da fase do sono e na predição da complexidade do sonho. Esses resultados sugerem que os relatos do REM possuem uma conectividade maior do que os relatos do N2 (i.e. as palavras recorrem com uma distância maior), o que, em nossa visão, está relacionado a diferenças subjacentes na complexidade dos sonhos. Esses achados também apontam para a análise de grafos como um método promissor no
campo dos sonhos, devido à sua relação com a complexidade do sonho e ao seu potencial de atuar como uma medida complementar ao comprimento do relato.

*Palavras-Chaves: Sonhos, sonhos não-REM, estrutura do sonho, comprimento do relato, análise de grafos.*
Introduction

The discovery of Rapid Eye Movement sleep (REM) (Aserinsky & Kleitman, 1953; Dement & Kleitman, 1957a) heralded the beginning of a new era of research in the area of sleep and dreaming. Using electroencephalography (EEG) to monitor participants who slept in a controlled laboratory setting, they observed cyclical physiological changes in subjects over the course of the night, such as variations in brain activity, muscle tone, body shifting and ocular movements. These changes across the sleep cycle have since been categorised into different stages, each with their own distinctive physiological markers. These include: the non-REM sleep stages (sleep onset -- N1, light non-REM -- N2, and deep non-REM/slow-wave sleep -- N3, formerly S3 and S4, see Rechtschaffen & Kales, 1968, Iber, Ancoli-Israel, Chesson & Chan, 2007 for an overview), and the REM state. The non-REM stages (N1, N2, N3) are characterized by, among other things, decreases in brain activity, responsiveness to external stimuli, body movements, body temperature and blood pressure, which drop to their lowest in N3. On the other hand, the REM state also known as "paradoxical sleep", is characterised by high levels of brain activity, complete muscle atonia in most of the body, a lack of thermoregulation, increased breathing, genital engorgement and, as the name indicates, brisk ocular movements. A full cycle through these states typically lasts up to 90 minutes, and changes throughout the night, where we tend to spend more time in N3 during the first half of the night and more time in REM during the second half (see Figure 1 for an illustration).
Figure 1. A hypnogram illustrating the distribution of sleep stages across the night. REM tends to increase in duration across the night while deep sleep (N3) tends to decrease. Image extracted from (URL: https://oneirology.co.uk/glossary-of-dreamy-terms/hypnogram/).

In addition to the abovementioned physiology, Kleitman and collaborators (Aserinsky & Kleitman, 1953; Dement & Kleitman, 1957a, 1957b) observed that awakenings during REM were highly associated with reports of dreaming (around 80%), compared to non-REM awakenings (around 10%). Given the almost perfect correlation between REM awakenings and dreaming, the few dream reports elicited from NON-REM were “best accounted for by assuming that the memory of the preceding dream persisted for an unusually long time” (Dement & Kleitman, 1957b). REM was thus proposed to be the exclusive physiological basis of dreaming (“REM = dreaming” perspective), a claim which later formed the basis for influential models of dreaming (e.g. Activation-Synthesis, McCarley & Hobson, 1977), and had considerable influence on the scientific field of dreaming.

In opposition to such a perspective, a later study by Foulkes (1962) demonstrated that by
using of a more inclusive definition of dreaming, termed *sleep mentation*¹, complex cognitive activity could be yielded in as many as 50% of NON-REM awakenings. Such NON-REM mentation could not be explained by remembered dreams from previous REM cycles, since awakenings from N1 (Foulkes, Spears & Symonds, 1966; Foulkes, 1962) and NON-REM naps (Foulkes & Vogel, 1965; Suzuki et al., 2004) reliably produced dream reports before the onset of the first REM cycle. Additionally, towards the end of the 20th century, another line of evidence emerged showing that patients with lesions to the mesolimbic-mesocortical dopamine system reported a complete cessation of dreaming, despite showing unaltered REM cycles (Solms, 1997, 2000). Collectively, these findings cast major doubt on the REM=dreaming perspective, and rather suggested that REM and dreaming are doubly dissociable states (Solms, 2000), each with distinct underlying neural mechanisms. While there is now a consensus amongst contemporary dream researchers that mentation can be recalled throughout the night in both REM and NON-REM stages, debate has now turned to investigating the nature of the differences present between dreaming in these sleep stages.

**Differences in REM and Non-REM Dreaming**

Possible differences in REM and non-REM are of particular interest to modern dream research, since they are informative of the underlying nature of the mechanisms that generate dreaming. Some researchers (e.g. Hobson, Pace-Schott & Stickgold, 2000) propose that REM and non-REM dreaming are generated by two separate underlying generators ("Two-gen models"), while others (e.g. Antrobus, 1983; Foulkes, 1985; Soms, 2000) propose that dreaming

---

¹ Unlike more traditional definitions of dreaming, sleep mentation refers to any mental experience that is experienced during sleep, such as thoughts, sensations and non-visual experiences that may not be considered to be genuine dreaming. While debate persists regarding definitions, here we use the term dreaming and sleep mentation interchangeably to refer to all mental experience that occurs during sleep.
is driven by one underlying generator ("One-gen") responsible for all of dream-like experiences during sleep. Proponents of the "two-gen" models cite evidence that demonstrates differences in the quality of dream reports obtained from REM and non-REM sleep stages and affirm that these differences are indicative of two distinct underlying sources for the production of dreams: one that produces the mentation experienced in REM and the other that produces the mentation experienced in non-REM. On the other hand, authors of "one gen" models claim that the these apparent differences in the quality in REM and non-REM are in fact quantitative in nature and reflect an artifact of underlying residual differences in report length. Since REM reports are typically longer than their non-REM counterparts (see section below), one-gen authors argue that the qualitative nature of REM and non-REM reports can only be meaningfully compared when residual differences in report length are partialled out. In this regard, a number of studies have found that many of the apparent differences that are reported tend to diminish and even disappear after statistical controls for report length are employed (Antrobus, 1983; Foulkes & Schmidt, 1983). However, even after utilising such controls, some differences persist (Casagrande, Violani, Lucidi, Buttinelli, & Bertini, 1996; Speth, Frenzel & Voss, 2013; Waterman et al., 1993). Furthermore, the partialling out of report length has been methodologically questioned, since it presupposes that it is the length of a report that causes dream quality and not the other way round (Hunt, Ruzycki-Hunt, Pariak, Belicki, 1993; Stickgold et al., 2000). By doing so, one may erroneously “dilute the measure of a defining characteristic of dreams” and may be “partialling a variable out of itself” (Hunt et al., 1993, p. 181). Such an interpretation is supported by Hunt et al.’s (1993) finding that one requires more words to describe a bizarre stimulus, as compared to a mundane one.

While an overall consensus is yet to be reached and debate persists in this regard,
additional investigations that probe the characteristic differences of REM and non-REM dream reports can help further our understanding of dreaming and the nature of the underlying mechanisms that generate them. In the following paragraphs, we present a summary of the literature that has been reported in relation to REM and non-REM dreaming differences, related to: recall rates, report length, dream quality and structural organisation. We will then consider how dreaming in both of these phases are affected by the time of night effect. This literature is in part derived from Stickgold et al. (2000) and Nielsen's (2000) extensive reviews based on studies conducted between 1953-2000. We also have included other more recent studies post-2000, which have supplemented and informed the current knowledge in the field (Cipolli et al., 2015; Fosse Stickgold, Hobson, 2004; McNamara, McLaren, Smith, Brown & Stickgold, 2005; Montangero & Cavallero, 2015; Nielsen, Kuiken, Hoffmann & Moffitt, 2001; Oudiette et al., 2012; Rosenlicht, Primich, McQuaid, Maloney, Feinberg, 2017; Speth, Frenzel & Voss, 2013; Smith et al., 2004; Speth, Harley, Speth, 2017; Stickgold, Malia, Fosse, Propper, & Hobson, 2001; Wamsley, Hirota, Tucker, Smith, & Antrobus., 2007).

**Recall rates**

The first distinction to be noted between REM and non-REM dreaming relates to recall rates, which led to the REM=dreaming controversy in the first place. An extensive review of 35 studies by Nielsen (2000) demonstrated that recall rates are considerably higher in REM (mean - 81.9% ± SD:9.0), compared to non-REM (43% ± SD 20.8). Another important observation was that the findings for REM dreams have been fairly consistent over the previous years ranging from 60-93% across the different studies. In contrast, there were major discrepancies in studies reporting non-REM recall, which ranged from 0-75% and showed a positive trend of increase for non-REM recall rates over time. This divergent trend in recall rates can best be explained by the
use of less restrictive operational definitions of dreaming as the years have gone by (Nielsen, 2000, see Figure 2). This is corroborated by Nielsen’s (2000) finding that the reported non-REM recall rates are inversely related to the stringency of the definition employed by the study. Such changes in definition tend to affect non-REM dreams to a greater extent, since unlike REM dreams, which tend to compose our most vivid and intense dreams, non-REM dreams are typically more thought-like and conceptual (Fosse et al., 2004), thus lacking some of the traditional properties that some may consider to be “genuine” dreaming. Despite such controversies, even when using a liberal definition, dreaming is consistently more frequent in REM compared to non-REM following controlled awakenings.

Figure 2. Graph showing changing recall rates in sleep mentation over the years. This is best explained by changes in criteria for the definition of ‘dreaming’, which became less stringent over the years. Extracted from Nielsen (2000).

Report length

The most robust difference found between REM and non-REM dreams in current literature relates to differing report lengths, where REM dreams have been consistently found to
be longer than non-REM ones. The most widely used measure to reflect report length is that of Total Recall Count (TRC, Antrobus, 1983), which was developed as an overall measure of information processing during sleep. TRC reflects the number of unique words present within a dream report, excluding repetitions, redundancies and external commentary not related to the dream content. Studies have found that REM reports are consistently longer than non-REM reports in terms of TRC (Antrobus, 1983; Antrobus, Kondo, Reinsel & Fein, 1995; Casagrande et al., 1996; Oudiette et al., 2012; Stickgold, Pace-Schott & Hobson, 1994; Stickgold et al., 2001; Wamsley et al., 2007, Waterman, 1993), raw number of words contained in the report (Casagrande et al., 1996; Foulkes & Rechtschaffen, 1964; Goodenough et al., 1965), as well as other measures, such as temporal units (Cavallero, Cicogna, Natale, Occhionero, & Zito, 1992; Foulkes & Schmidt, 1983; cf. Cicogna, Natale, Occhionero, & Bosinelli, 1998) and subjectively judged dream length (Foulkes 1962).

**Dream quality**

Studies indicate that REM and non-REM dreams are qualitatively different in a number of important respects. Firstly, REM dreams are typically rated as more intense, complex (Wamsley et al, 2007; Oudiette et al., 2012) and visually hallucinatory (Fosse et al., 2004), compared to non-REM dreams, which are typically rated as more thought-like (Wamsley et al., 2007; Fosse et al., 2004) and conceptual (Foulkes & Rechtschaffen, 1964). Secondly, dreams from REM tend to be more bizarre than those from non-REM, as rated by judges (Goodenough, Lewis, Shapiro, & Sleser, 1965; Ogilvie, Hunt, Sawicki, & Samahalskyi, 1982) and by the dreamers themselves (Foulkes, 1962; Foulkes & Rechtschaffen, 1964; Rechtschaffen, Verdone, & Wheaton, 1963). They also contain more bizarre elements (Oudiette et al., 2012; Wamsley et

---

2 Cicogna et al. 1998 found no difference in the number of temporal units when comparing spontaneous morning awakenings in N2 and REM.
al., 2007), in the form of incongruous and discontinuous elements. Thirdly, REM dreams are more perceptually vivid, as evidenced by subjective (Rechtschaffen et al., 1963; Antrobus et al., 1995; Wamsley et al., 2007) or external ratings (Casagrande et al., 1996; Ogilvie et al., 1982). They also tend to contain more visual imagery (Goodenough et al., 1965; Foulkes, 1962), which has been subjectively rated as brighter and clearer (Antrobus et al., 1995; Wamsley et al., 2007), and more auditory verbal experiences (Antrobus et al., 1995; Speth et al., 2013; Speth et al., 2017). Fourthly, REM dreams are more emotional, both in terms of their overall affective intensity (Smith et al., 2004; Wamsley et al., 2007) and the relative prevalence of dream emotions (Cavallero et al., 1992; Rechtschaffen et al., 1963; cf. Cicogna et al. 1998), while they are also more likely to contain aggressive interactions (McNamara et al., 2005) and motivational content (Smith et al., 2004). Lastly, REM dreams tend to be more kinesthetically engaging, in terms of the judged presence of activity (Goodenough et al., 1965; Foulkes, 1962; Foulkes & Rechtschaffen, 1964) and through the linguistic analysis of dreamer activity (Speth & Speth, 2017).

**Structural organisation and narrative complexity**

A final line of evidence to consider comes from studies comparing REM and non-REM dream reports in terms of their structure, narrative complexity and story-like organisation. Nielsen, Kuiken, Moffitt, Hoffmann, and Newell (1983) found that dream reports collected after REM displayed more of a story-like organisation compared to reports collected after N2. Similarly, by comparing the basic attributes of story narratives, Nielsen et al. (2001) found that REM dream reports more often possessed at least one story constituent and episodic progression compared to N2 dreams. On the other, Cicogna et al. (1998) found no difference in the narrative continuity of REM and N2 dream reports obtained from spontaneous morning awakenings;
similarly, by using a subsample from this same study (Cicogna et al., 1998), Montangero and Cavallero (2015) found no differences in a microanalysis of 14 dream reports matched for report length.

**Time of night effect**

While the differences outlined above point to overall between-stage differences in dreaming, another important factor that may influence dreaming, is the time of night in which the dream occurs. Over the course of a typical night, circadian cortical activation tends to increase, which is associated with characteristic changes in dreaming. Some of these time-dependent changes appear to be common to both sleep phases. For example, both REM and non-REM dream reports become longer (Casagrande et al., 1996; Pivik & Foulkes, 1964; Stickgold et al., 2001), more dreamlike (Pivik & Foulkes, 1968; Rosenlicht et al., 2017) hallucinatory (Fosse et al., 2004) and bizarre (Antrobus et al., 1995; Wamsley et al., 2007), while related increases are observed in verbal and visual imagery, whose appearance becomes clearer and brighter towards the late-morning (Antrobus et al., 1995; Wamsley et al., 2007). However, some of these effects appear to be sleep stage specific, where, for example, selective increases in emotionality are seen in REM dreaming (Wamsely et al., 2007), while a selective decrease in directed thought has been observed in non-REM dreaming (Fosse et al., 2004). Additionally, the narrative complexity of REM dreams increases across the night (Cipolli, Bolzani, & Tuozzi, 1998; Cipolli et al., 2015), although such related changes in non-REM dreaming are yet to be investigated.

**Summary**

The findings cited above suggest that REM and non-REM dreaming differ in a number of important respects. Specifically, REM dreams are longer, more frequent, dreamlike, complex, perceptually vivid, bizarre, emotional, kinesthetically engaging and story-like compared to non-
REM dreams. These differences appear to be influenced by the time of night in which the report was obtained whereby dream intensity and its related qualities tend to increase. Questions remain over the extent to which the observed differences are a reflection of intrinsic qualitatively different processes or residual differences of quantitative differences in report length. Further research investigating differences in REM and non-REM dreaming in relation to report length can help supplement current literature. While previous studies have analysed the narrative complexity and story-like nature of dream reports, to our knowledge, the word-by-word structural organisation of REM and non-REM dream reports is yet to be investigated and meaningfully compared. One suitable method for such an evaluation is through the analysis of speech graph attributes (SGA).

**Analysis of Speech Graph Attributes (SGA)**

SGA, like other mathematical graph-based techniques, derives from graph theory where networks \((G = N, E)\) are defined by a given number of vertices/nodes \((N = \{1,2,3\ldots\})\) and a set of edges \((E = 1,2,3\ldots\})\), which maintain connections or relations between them. These graph relations can either be directed or undirected depending on the nature of the relationship that is being modelled. For example, one’s family tree may be characterized through representing each node as a family member, and direct bloodline relations through edges; these familial relations could either be represented by undirected edges (Mary – John), indicating kinship; or directional ones, showing the track of parent-child relations in successive generations (Mary → John). Edges may also either be weighted if the particular relationship being represented is of varying degrees (e.g. Mary is more related to her father than to her uncle) or unweighted, if a binary absent/present relationship exists (e.g. John is either Mary’s child or he isn’t). Evidently, as the above example illustrates, the same phenomenon may be represented in a variety of ways; as a
result, care is needed in order to select the appropriate network features that best represent the problem to be investigated. Indeed, representing a given phenomenon as a network is a “theoretical act”, which “commits one to assumptions about what is interacting, the nature of that interaction, and the time scale on which that interaction takes place” (Butts, 2009, p. 416).

In the case of SGA, each node is represented by a lexeme (Mota et al., 2012) or an uttered word (Mota Furtado, Maia, Copelli & Ribeiro, 2014; Mota et al. 2016a; Mota, Copelli, Ribeiro 2017), while the succession of one’s speech is represented by directed, unweighted edges (see Figure 3). Given that one may utter the same sentence more than once or the same word repeatedly, SGA also permits multigraphs – graphs in which there may be two or more edges between two corresponding nodes, as well as loops where a node can connect directly back to itself. Once a speech graph is defined, several attributes can be calculated, leading to a detailed quantitative profile of its structure. Importantly, SGA is not an overall measure of semantics, and thus is not informative of content; rather it is a measure of speech structure, which can reflect the ways in which people organise their thoughts. In this way, SGA is “based not on what is said, but on how it is said” (Mota et al., 2012, p. 1).

![Figure 3](image-url). Illustration of graph representations of speech, where each word is represented by a node and the succession of speech is represented by directed edges. Extracted from Mota et al., 2014.
In recent years, SGA has been used in a variety of interdisciplinary applications, from education and literature to computational psychiatry and developmental psychology. For example, differences in graph connectedness reveal a striking resemblance of the age-related changes in the organisation of children’s memories and in the historical development of literature (Mota et al., 2016b). Graph connectedness can also predict cognitive functioning and reading ability in 6-8 year olds (Mota et al., 2016c), and distinguish between Alzheimer’s, mild cognitive impairments and matched controls in the elderly (Bertola et al., 2014). In perhaps its most fruitful application to date, SGA has been used as a “privileged measuring lens into thought” (Mota et al., 2012, p.1) by discerning differences in graph connectedness (Mota et al., 2014) and the graph’s random-like nature (Mota et al., 2017) by comparing verbal reports of controls and patients suffering from schizophrenia. Of particular relevance to the present work is that out of a variety of verbal memory reports, most recent dream reports appear to be especially informative of underlying thought disturbances in psychosis (Mota et al., 2014), and particularly of the negative symptoms of schizophrenia (Mota et al., 2017). Further studies of dream reports using SGA may 1) shed light on the inherent structural differences of dreams from different sleep stages, 2) elucidate the special status of dreams in revealing underlying differences in thought disturbances, and 3) enhance our understanding of SGA as a method for psychology research.
Aims and Objectives

In the present study, we aimed to investigate the structural organisation of REM and N2 dream reports through applying SGA to a previously collected sample of dream reports obtained from controlled laboratory awakenings across the night. The research had two main objectives: firstly, to investigate whether REM and non-REM reports are differentially structured in terms of graph connectedness and graph random-likness; and secondly, to evaluate how graph connectedness compares to one of the most widely used measures in dreaming, TRC in predicting differences in sleep stage and dream complexity, as measured by the Perception Interaction Rating Scale (PIRS). In this regard, four specific hypotheses were investigated:

Firstly, in order to test whether our sample conforms to previous studies that have found longer reports in REM, we tested the hypothesis that REM reports are longer than N2 ones.

**Hypothesis 1**: REM reports will be significantly longer than non-REM reports in terms of report length (TRC).

Secondly, we aimed to evaluate whether intrinsic structural differences exist in REM and non-REM dream reports transformed to word graphs, by analysing and comparing their graph connectedness and random-like quality.

**Hypothesis 2**: REM reports will be structurally different to non-REM ones in terms of graph connectedness and their approximation to random graphs.

Thirdly, we aimed to evaluate whether there is a time of night effect present within these variables, whereby REM and non-REM dreams change across the night in terms of TRC and graph structure:

**Hypothesis 3**: Graph Structure and TRC will change as a factor of the time of night.

Finally, we aimed to evaluate how graph connectedness relates to TRC in discerning
sleep stage and predicting overall dream complexity.

**Hypothesis 4:** Graph structure and TRC will be able discern which sleep stage a dream report is obtained from.

**Hypothesis 5:** Graph structure and TRC can predict differences in the external ratings of dream complexity, as measured by the Perception Interaction Rating Scale (PIRS).
Methods

The data analysed in the present study is derived from a Master’s thesis, originally collected at the University of Cape Town (see Wainstein, 2013, for full overview of study). The initial investigation explored non-REM dreaming in relation to sleep microstructure and the cycling alternating pattern. While non-REM dreaming was the main focus of the study, both REM and non-REM dream reports were collected from participants and thus the data was suitable for the testing of the hypotheses for the present study. The research used a quasi-experimental repeated measures design whereby participants spent two to three inconsecutive nights in a sleep laboratory to provide dream reports. Researcher, Danyal Wainstein (DW), the Master's student involved in the research and co-author of this paper, conducted the controlled awakenings and dream report collection. External judges blind to the original aims of the research were involved in the transcription of dream reports and the rating of dream reports.

Participants

Twenty-two subjects (14 female) aged 18-25 were recruited from the University of Cape Town to participate in the sleep study. All participants were undergraduate Psychology students of the university and were recruited via an online questionnaire. This questionnaire acted as an initial screening measure, which was followed by an-person interview at the university. Subjects who met the suitable criteria were subsequently invited to participate in the sleep study. Those who participated in the sleep study were compensated financially for their participation (see section on Ethics and Confidentiality). Potential participants were evaluated according to the following inclusion/exclusion criteria:

Inclusion criteria

*Verbal fluency:* Subjects were required to all be fluent English-speakers, assessed by a
minimum score of 100 for verbal IQ on the *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999). This was done to minimise any individual discrepancies in proficiency that may affect the ways that dreams are reported. Such a consideration is especially important in the case of SGA where graphs, a proxy for speech structure, are liable to be affected by linguistic proficiency.

**Sleep quality:** Subjects were included who reported adequate sleeping habits and received a score of 5 or lower on the *Pittsburg Sleep Quality Index* (PSQI; Buysse, Reynolds, Monk, Berman, & Kupfer, 1989)). This was done to avoid abnormal participant sleeping patterns during the study, which may affect the results.

**Subjective dream recall:** Only respondents that reported to be moderate to frequent dreamers (at least once every two weeks or more) were included in the study, as measured by Schredl’s self-reported dreaming scale (Schredl, 2004). This was performed to reduce inter-participant differences in habitual dream recall and include participants who would reliably produce dream reports.

**Exclusion criteria**

**Substance-use:** Since recreational drug use (frequent or occasional) may affect aspects of sleep and dreaming (see Schierenbeck, Riemann, Berger, & Hornyak, 2008, for a review), participants were excluded who reported the use of illicit substances. Additionally, smokers were excluded due to the influence of nicotine on the sleep cycle (Page, Coleman, Conduit, 2005).

**Psychiatric or sleep disorders:** The *Mini International Psychiatric Inventory* (M.I.N.I; Lecrubier et al., 1997) evaluated whether participants had any persistent psychiatric disorders, while the PSQI evaluated any existing sleep disorders. Participants were excluded according to any disorder present to control for their effects on sleep and dreaming.
Procedure

Sleep study

The sleep study took place at a hospital sleep laboratory in association with the university where participants spent 3-4 inconsecutive nights, consisting of one adaptation night, followed by 2-3 experimental nights. The adaptation night served in order for participants to familiarise themselves with the laboratory setting; thus, no controlled awakenings or sleep recordings were conducted. On experimental nights, the sleep of participants was monitored by polysomnography (PSG) and controlled awakenings were performed in order to obtain dream reports and related questionnaire data. Each experimental night was separated by 2-7 days. This helped minimise any sleep deprivation effects that may result from the experimental awakenings.

Experimental night protocol

Participants were asked to refrain from using any alcohol on the day of the experiment and not to use caffeine after 15:00. On the experimental nights, participants arrived at around 19:00 and were prepared for sleep monitoring. DW switched off the lights at 22:00 and woke the participants up at 6:00, totalling approximately 8 hours of sleep recordings per session. Participants were woken for the collection of dream reports up 5-6 times over the course of the night, including the morning awakening. In the morning, participants were thanked and compensated accordingly.

Awakening protocol

Controlled awakenings were performed in REM, N2 and N3 sleep stages according to the online presence of defining PSG characteristics for the respective stages. For REM, the controlled awakenings were conducted after 5-10 minutes of the presence of muscle atonia (via EMG), desynchronised “saw-tooth” waves in brain activity (via EEG) and distinct jagged eye-
movements (via EOG). For N2 awakenings, the presence of theta waves, sleep spindles and K-complexes (via EEG) was used as defining criteria, while N3 consisted of the presence of synchronised, high-amplitude delta waves and diminished muscle tonus (via EMG). In the case of N2 and N3, the length of time spent in a specific sleep stage was not always the same prior to the awakening, since sequences of sleep stability/instability were difficult to predict. Following any given controlled awakening, participants were allowed to go back to sleep for at least 40 minutes of uninterrupted sleep before a subsequent awakening. Additionally, at least 15 minutes were allowed to pass after a period of REM.

**Dream report collection**

When a participant met the defining PSG criteria for the desired stage of sleep, the researcher (DW) entered the room where the subject was sleeping and called out their name until they verbally indicated that they were awake. DW then asked the subject to recall and report all dream contents that he/she could remember. The dialogue between participant and DW was based on the protocol established by Foulkes, Spear & Symonds (1966) and Antrobus et al. (1995), according to the following general interaction:

“Tell me everything that was going through your mind just before you were awoken.”

After the subject finished his/her report, the researcher asked:

“Is that everything that you can remember?”

Whenever it was unclear whether the dream was visual, the researcher asked:

“Could you see what was going on? Was it visual?”

Following collection of the verbal dream report, participants were asked to fill out a questionnaire containing a number of Likert scales pertinent to the aims of the original study. Verbal dream reports were recorded using a voice recorder and later transcribed by an external
judge blind to the conditions of the respective awakenings.

Materials

Speech Graph Analysis (SGA)

Software was used to convert transcribed speech into directed graphs, where nodes represent words and directed edges represent word succession. Speechgraphs is a custom-made java software developed by investigators at the Federal University of Rio Grande do Norte for the graph-theoretical analysis of speech (available for free at -- http://neuro.ufrn.br/softwares/speechgraphs). From the resultant transformation, a total of 14 SGA are yielded relating to lexical diversity: Nodes; connectedness: Edges, Largest Connected Component (LCC), Largest Strongly Connected Component (LSC); recurrence: Repeated Edges (RE), Parallel Edges (PE), L1 (loops of one), L2 (loops of two), L3 (loops of three); and network size: Average Total Degree (ATD), Density, Diameter, Average Shortest Path (ASP), Average Clustering Coefficient (CC). Once the transformation is complete, a dream report, or section of a dream report, may be represented according to one or more subgraphs. Distinct subgraphs are made possible by separate paragraphs in the report. If the new paragraph contains words that were previously mentioned in the transformation, then it will be connected to the respective subgraph and will not be distinct; if no word is present, the subgraph will be distinct (see Figure 4 for a graphical illustration).

While there are a number of graph measures derived from SGA, in the present study we selectively chose to evaluate attributes of graph connectedness and graph random-likeness (see Figure 4), which have been informative of underlying changes in thought, such as those in schizophrenia (Mota et al., 2014; Mota et al., 2017). Graph connectedness can be seen to reflect words recurring over a longer or shorter range: long range recurrence is reflected in larger graph
connectedness scores.

**Attributes of graph connectedness**

1. Edges (calculated by the total number of edges present in the graph)
2. Largest Connected Component (LCC; calculated by the number of nodes in the maximal subgraph in which all nodes are connected to one another)
3. Largest Strongly Connected Component (LSC; calculated by the number of nodes in the maximal subgraph in which all nodes are mutually accessible to one another, i.e. A leads to B, B leads to A).
Original Report

“Er I was at school, part of some kind of test, separating out people that were smart and a couple of us starting like boycotting it and striking against it. Yeah that’s all I’ve got.

Mmm... no.”

Figure 4. Visual Representation of a dream report represented as a graph. Nodes indicated in red, edges indicated as black arrows. The largest connected component (LCC) has 31 nodes.
Controlling for report length: the sliding window method

Given that (1) connectedness attributes are highly collinear with word count (Mota et al., 2014) and (2) REM reports are typically longer than those of non-REM (e.g. Antrobus, 1983), any overall connectedness differences found when using the entire reports in the transformation would be heavily confounded by differences in report length and thus would not be informative. In order to control for such residual effects, we employed a sliding window method, which controls for word count by dividing the report up according to the window size employed. Such a method has been used in order to meaningfully analyse and compare the speech of people suffering from psychosis, who generally are less verbose compared to matched healthy controls (Mota et al., 2014). The method works by moving a window of a fixed word length along the text document, calculating separate attribute scores for each respective window (see Appendix A, for an illustrated example). After reaching the end of the document, the mean value for each attribute is calculated \( \text{SGA} = \sum \text{Attributes} / \text{No. of windows} \) from the respective windows and is returned to the user as an output. Depending on the length of document and precision required, one can adjust the length of the window and the overlap between windows according to one’s needs. The higher the overlap between successive windows, the higher the resolution obtained from the resultant transformation. For the present study, we utilised a window of 30 words with an overlap of 29 words. The window size was chosen based on findings that show 30-word windows to be more informative than comparatively smaller sized windows (10/20 words) in previous studies (Mota et al., 2014; Mota et al., 2017). A bigger window (e.g. 50 words) was also not considered since many reports would be excluded for not meeting the minimum number of words needed for a single window. Given that computing power was not an issue in our dataset, we applied a window with maximum
Calculating random graphs: comparing random-likeness of reports

In order to investigate the random-like connectedness of dream reports, we compared each transformed report to 1000 random graphs, which are assembled using the same number of nodes and edges, but whose word-order is arbitrarily shuffled (see Appendix B for an illustration). Using a distribution of 1000 random graphs for each respective dream report, we divided the original LCC and LSC values for the entire report by the mean of the random LCC and LSC distributions in order to obtain an estimate of random-likeness (see Appendix B for formula). In this sense, graphs that are more random-like are those who approximate more to the calculated mean of the random graph distribution. Since original graph scores are divided by random graph scores, those that approximate to 1 are more random-like. While Edges represent a measure of graph connectedness, their total number is not affected by shuffling word order and are thus not included as a measure of random-likeness. Since graphs are compared to their respective random graph distributions where nodes and edges are maintained constant, differences in report length are suitably controlled for in the LCC and LSC random-like comparisons.

Total Recall Count (TRC)

TRC is an objective measure of report length usually rated by two or more external judges. It is measured by the total number of words used to describe any mentation experienced prior to awakening, excluding repetitions, redundancies, “ums” and “ahs”, corrections and external commentary on the dream (Antrobus, 1983). It is widely used in dream research and known to be one of the best measures to distinguish REM and non-REM mentation (Antrobus, 1983). The measure has been more recently revised by Wamsley et al. (2007) under the new
name *Word Count Index*.

**Perception-Interaction Rating Scale (PIRS)**

The PIRS was constructed for the purposes of the original study (Wainstein, 2013) and was developed as a measure of overall dream complexity. The scale is rated by two or more external judges. They are trained to score the dream reports according to an ordinal scale from 0-9, according to the level of complexity described within the report. In this sense, PIRS is similar to other ordinal scales that code dream reports according to their relative complexity (e.g. Orlinsky’s classification measured on a 0-7 scale, Oudiette et al., 2012). Here, we will briefly describe the various levels of the scale that, divided into its five broad categories:

- **(0) “No Recall”**. A score of 0 refers to a situation in which the participant reports having experienced no mental experiences.
- **(1) “White Dream”**. A score of 1 is given to a situation in which the participant feels as if they were dreaming but cannot remember any details of the dream.
- **(2-3) “Nonvisual Recall (Conceptual)”**. A rating of 2-3 refers to non-visual experiences that typically involve conceptual or thought-like mentation. A score of 2 is given to simple, fragmented experiences that are isolated and lack coherence, while a score of 3 is given to more complex, ongoing coherent experiences, which may be accompanied by non-visual perceptions such as sounds (voices or music), smells or kinesthetic sensations.
- **(4-5) “Visual/Perceptual, Non-Narrative/Incoherent”**. A rating of 4-5 includes experiences with visual perception that take place within an isolated scene/incident that is unrelated to an ongoing narrative. A rating of 4 is given to the most basic of visual experiences that do not involve any interactions, while a rating of 5 is scored for visual experience with an isolated interaction with other characters or with the dream.
environment.

- (6-9) “Visual/Perceptual Dreaming, Part of an Ongoing Narrative”. A rating of 6-9 refers to experiences that take place within a visual/perceptual dream environment and describe one or more interactions that are related to an ongoing series of events or a story/narrative. The numbers increase in accordance with the complexity of the narrative described in the report, relating to the number of stages of development (6 - one, 7 - one or two, 8 - three or four, 9 - five or more) and the intensity of the interactions present (6 - limited interaction, 7 - moderate interaction, 8 - prolonged interaction, 9 - intensive interaction).

**Polysomnography (PSG)**

In order to monitor sleep, a NeuroFax EEG9000 polysomnograph was used in conjunction with the software Polysmith Online Version 6. PSG measures and records levels of brain activity (EEG), heart rate (electrocardiography, ECG) muscle tone (electromyography, EMG) ocular movements (electrooculography, EOG). In conjunction, these measures were used to assist in the classification of sleep stages for the controlled awakenings, which was conducted both online and offline.

**Voice recorder**

A portable voice recorder was used to record the verbal mentation reports of the participants immediately after being awoken. These reports were then saved as audio files and later transcribed by an external judge.

**Ethics and Informed Consent**

The study was approved by the Psychology department’s ethics committee (see Appendix C) prior to data collection. Subjects were compensated financially in accordance with their
participation where they received R400 (≈ 100 R$) for spending two experimental nights in the sleep laboratory (R200 per experimental night). All participants were fully informed about their participation in the study and signed consent forms. Participant information was kept strictly confidential. Participants were free to withdraw from the study at any time should they wish, however they would only be compensated financially for the number of experimental nights spent in the sleep laboratory. The study and the compensation of participants was conducted in accordance with the established guidelines set out by the University of Cape Town’s Code for Research and the Helsinki Declaration for human experimentation.

**Data Analyses**

**Wilcoxon sign-rank tests**

In order to test for normality, Shapiro-Wilk tests were carried out for TRC and the connectedness attributes (Edges, LCC, LSC, LCCR, LCSR) in both REM and N2 conditions. Several tests were found to be significant (all p < .001, see Appendix D) indicating non-normal distributions within the data. Given that dream reports from the same individual constitute independent samples, and participant medians tend to vary (see Appendix D) we utilised a nested design comparing participant medians from the two respective stages. Although our sample size is reduced and some data points are potentially lost, such a design controls for the overrepresentation of participants who contributed many dream reports and takes into account individual differences in TRC and speech style.

**Generalised linear models**

In order to see whether graph connectedness can distinguish between sleep stages and how it relates to TRC in this discernment, we implemented generalised linear models with a binomial outcome (REM vs. N2) using the native ‘glm’ command in the R environment (R Core
Team, 2017). Also known as a logistic regression, such a method was chosen to complement the Wilcoxon sign-rank tests, as it allows one to test whether two measures are complementary in predicting a given binomial outcome. In our case, since we are interested in whether graph attributes can supplement a widely used existing measure (TRC), we can use a hierarchical method in order to construct and compare models in order to evaluate whether the inclusion of a particular variable can significantly improve the fit of a model. Such a method also possesses the advantage of being able to utilise the full data set in the analysis.

**Cumulative link models**

In order to estimate the relationship between graph connectedness, time of night (i.e. order of awakenings) and external ratings of dream complexity (i.e. PIRS), we employed cumulative link models using the ‘clm’ and ‘clm2’ commands from the Ordinal package (Christensen, 2015) in the R environment (R Core Team, 2017). This model is otherwise known as an ordinal regression or proportional odds model, and allows for the prediction of an ordered variable where the distance between respective points of an ordered scale cannot be assumed to equal (i.e. $1 - 2 \neq 2 - 3 \neq 3 - 4$ etc.). Given that PIRS is measured along an ordinal scale, this model was preferred over the linear model alternatives. We also used this in order to model the order of awakenings conducted, since the distance between successive awakenings cannot be assumed to be equal in our data. While ideally we would use the exact time of night in order to evaluate the effect of the circadian rhythm on dreaming, given that the data was not available at this stage, we used the order of awakening as a statistical compromise. As with the generalised linear models, we used a hierarchical method in order to evaluate the unique contribution of graph attributes towards improving the fit of models containing theoretically pre-established variables (sleep stage and TRC).
Significance values and goodness of fit

A traditional threshold of 5% for was used for the general testing of statistical significance, with the inclusion of a Bonferroni control for multiple significance tests (see below). For the Wilcoxon sign-rank tests, significance was calculated using the exact method. Such a method is preferable in smaller samples, which is the case here (n = 38). For the cumulative link and the generalised linear models, in order to calculate significance, we compared the log likelihood ratios ($\chi^2$) of models with and without the predictor variable that is being tested. Significant results indicate that the overall fit of a model has been significantly improved by the addition of a given variable. Values of Akaike’s Information Criterion (AIC) were also included in order to evaluate model quality. AIC is a relative measure of the overall goodness of fit of a model, where model simplicity is also taken into account. Lower AIC values indicate that a better fitting model; however, such values are not absolute and can only be interpreted relative to other models fitting the same data.

Controlling for family-wise error rate

A key issue in statistical hypothesis testing is the balancing of Type I “false positive” and Type II “false negative” errors. The use of multiple tests of significance is known to artificially inflate the chance of erroneously finding a significant result. Since our analyses compose multiple tests, it is important to consider a suitable method to ensure we avoid mistakenly finding significant results. Perhaps the most widely used measure to be employed in this regard is the Bonferroni method, which calculates a significance alpha by dividing the prescribed p-value (usually 0.05 in most fields) by the number of statistical tests used. However, such a method has been criticised for being too stringent and can conversely inflate the chances of incorrectly assuming that a particular result is not significant (Perneger, 1998). Another factor for
consideration is that the Bonferroni adjustment assumes that the significance of each test is independent, which often is not the case. In the case of our analyses, we tested a total six measures: three related to graph connectedness (Edges, LCC, LSC), two related to graph random-likeness (LCCr, LSCr) and one related to report length (TRC). Given that connectedness and random-like measures are highly correlated to one another (see Appendix D), in order to balance Type I and Type II errors, we chose to group measures according to their theoretical category, reflecting in three distinct groups: graph connectedness, graph random-likeness and report length. As a result, we used a Bonferroni correction for three tests, resulting in an adjusted significance alpha of 1.67% ($\alpha < .0167$).

**Effect sizes**

For the Wilcoxon sign-rank tests, effect sizes were calculated using the R statistic developed by Rosenthal (1994), which is calculated by dividing the Z-score over the square-root of the sample size ($r = Z/\sqrt{N}$). For cumulative link models and generalised linear models, overall estimates of effect sizes were calculated using the NagelKerke function in the ‘rcompanion’ package (Mangiafico, 2017), which provides a pseudo-R-squared estimate for both the generalised linear models and cumulative link models.

**Random effects and independence of observations**

The independence of observations assumption refers to the fact that data points (or residuals in the case of models) should not dependent on one another (i.e. the value or residual of a given data point should not affect the value or residual of another). Since our data consists of multiple dream reports coming from different participants across a number of experimental nights, such an assumption violation needs to be considered for the present analyses. A visual exploration of the data tends to show inter-subject differences in speech (see Supplementary
Material), although differences across the experimental nights were less clear. In order to control for this assumption, in the Wilcoxon sign-rank tests, we used the median scores for each participant in REM and N2 in order to control for possible observation dependence and the overrepresentation of certain participants who contributed more dream reports to the analysis. Each participant thus contributed equally towards the sample. For the hierarchical model analyses, we considered the respective mixed model variants (generalised linear mixed model and cumulative link mixed model), which allow for the consideration of random effects. Accordingly, we entered the participant IDs and the experimental night number as nested random effects. Such an approach ensures that any significant residual differences in participants over the respective nights are controlled for by the respective models. However, we found that the inclusion of these random effects did not improve the fit of the respective models and did not affect the model outcomes in anyway. Additionally, their inclusion resulted in a few convergence issues. As a result, we chose to implement the simpler GLM and CLM variants containing only fixed effects. Given that the results were not altered from including random effects, we can be confident that the results obtained here have not been contaminated by a violation of the independence assumption.
Results

Descriptive Statistics

Participants

Twenty-two undergraduate students (14 females; 8 males) took part in the sleep study (Age: Mean = 19.71 ± 1.59). All were fluent speakers of the English language (WASI score > 100: Mean = 115.95 ± 7.97), reported good sleeping habits (PSQI score < 5: Mean = 3.38 ± 1.53), to be moderate-to-frequent dreamers and were free of any psychiatric/medical disorders that could potentially affect one’s sleep and ability to report dreaming mentation (as assessed by M.I.N.I.). Of the 22 participants, three were excluded from the final sample, due to poor sleep architecture (one), extreme sleep inertia (one), and missing data (one). As a result, dream reports obtained from 19 participants (14 females; 5 males) were included in the final data analysis.

Dream recall

A total of 185 controlled awakenings were performed during REM and N2 sleep, resulting in the collection of 135 dream mentation reports from 19 participants (see Table 1 and Figure 5, for an overview). Dream recall was more prevalent in REM, while white dreams and no recall were more prevalent during N2. The recall rates found (REM - 90.7%, N2 - 65.6%) are located just outside the upper bounds of Nielsen’s (2000) recall estimates (REM - 81.9% ± SD 9.0; non-REM: 43% ± SD 20.8), but are nonetheless consistent with studies that have employed more liberal definition of dreaming (e.g. Foulkes, 1962), as was employed here. The elevated proportion of dreams in N2 (86) compared to REM (49) is a reflection of the greater number of awakenings that were performed in N2, since non-REM dreaming was the main interest of the original protocol (Wainstein, 2013). For the final sample, 10 dream reports (3 – REM, 7 – N2) were excluded as they did not meet the minimum word count of 30 words needed for graph
transformation.

**Dream complexity**

The 125 dream reports utilized in the final sample were rated according to the PIRS (2-8) that measured the overall quantity and quality the dream described. Reports describing conceptual, non-visual experiences were more prevalent in N2, while dreams that contained an ongoing narrative were more prevalent in REM. These are in line with previous studies indicating that, typically, REM dreams are more hallucinatory (Fosse et al., 2004), intense (Wamsley et al., 2007) and story-like (Foulkes & Schmidt, 1983); they also agree with previous studies showing that non-REM dreams are comparatively more thought-like (Fosse et al., 2004) and conceptual (Foulkes & Rechtschaffen, 1964).
Table 1

*Awakenings across the REM and N2 sleep stages (original sample, n = 185)*

<table>
<thead>
<tr>
<th></th>
<th>REM</th>
<th>N2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dream Report</td>
<td>49 (90.7%)</td>
<td>86 (65.6%)</td>
<td>135 (62.9%)</td>
</tr>
<tr>
<td>White Dream</td>
<td>1 (1.9%)</td>
<td>21 (16.0%)</td>
<td>22 (18.4%)</td>
</tr>
<tr>
<td>No Recall</td>
<td>4 (7.4%)</td>
<td>24 (18.3%)</td>
<td>28 (18.7%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>54 (100%)</strong></td>
<td><strong>131 (100%)</strong></td>
<td><strong>185 (100%)</strong></td>
</tr>
</tbody>
</table>

*Note*: Numbers are represented by frequencies; their respective prevalence is quoted in parentheses. A white dream refers to an experience where someone feels as if they were dreaming but are unable to recall any of its contents.

Table 2

*Dream complexity in REM and N2 (final sample, n = 125)*

<table>
<thead>
<tr>
<th></th>
<th>REM</th>
<th>N2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonvisual Recall</td>
<td>3 (6.5%)</td>
<td>19 (24.0%)</td>
<td>22 (62.9%)</td>
</tr>
<tr>
<td>Isolated Visual Imagery</td>
<td>23 (50.0%)</td>
<td>50 (63.3%)</td>
<td>73 (18.4%)</td>
</tr>
<tr>
<td>Part of Ongoing Narrative</td>
<td>20 (43.5%)</td>
<td>10 (12.7%)</td>
<td>30 (18.7%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>46 (63.2%)</strong></td>
<td><strong>79 (36.8%)</strong></td>
<td><strong>125 (100%)</strong></td>
</tr>
</tbody>
</table>

*Note*: Numbers are represented by frequencies; their respective prevalence is quoted in parentheses.
Figure 5. Bar-graphs showing distribution of PIRS scores in mentation reports collected after REM and N2 awakenings. The bar graph on the left shows awakenings that resulted in a score of 0 (no recall), 1 (white dream), or > 1 (dream report). The second bar graph shows, of those who were rated as dream reports, whether the dream report was related to nonvisual recall (2-3), isolated visual imagery (4-5) or part of an ongoing narrative (5-9). The y-value refers to the respective proportion of dreams as a percentage.
Inferential Statistics

Differences in graph structure and TRC

For hypothesis testing, we first aimed to investigate the possible differences between REM and N2 reports in accordance with our first and second hypotheses. We used two-tailed Wilcoxon sign-rank tests to compare the participant medians obtained in REM and N2 sleep (see Table 3 and Figure 6 for an overview of results). The results found that REM reports had significantly higher LCC and TRC scores compared to N2, a difference reflected with a large effect size. While differences were found in Edges and LSC in the same direction, such differences were not statistically significant according to our adjusted significance alpha ($\alpha < .0167$). Nonetheless, they show a significant trend in the same direction where REM graphs show larger connectedness. Finally, no significant differences were found in how random-like REM and N2, in terms of LCCr and LSCr. As a result, our first hypothesis that REM reports are longer than N2 reports was corroborated, while our second hypothesis that REM and N2 graphs are differentially structured was only partially corroborated, since differences were only found in graph connectedness and not in their structural random-likeness.
Table 3

Table showing results from Wilcoxon sign-rank tests (n = 38).

<table>
<thead>
<tr>
<th></th>
<th>N2</th>
<th>REM</th>
<th>Z-score</th>
<th>effect size (r)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRC</td>
<td>35.00 ± 13.28</td>
<td>49.00 ± 40.98</td>
<td>-3.04</td>
<td>.492</td>
<td>.002</td>
</tr>
<tr>
<td>Edges</td>
<td>28.38 ± 0.46</td>
<td>28.69 ± 0.52</td>
<td>-2.05</td>
<td>.333</td>
<td>.040</td>
</tr>
<tr>
<td>LCC</td>
<td>22.50 ± 1.16</td>
<td>23.84 ± 0.09</td>
<td>-2.05</td>
<td>.509</td>
<td>.002</td>
</tr>
<tr>
<td>LSC</td>
<td>15.99 ± 2.01</td>
<td>17.67 ± 2.62</td>
<td>-3.14</td>
<td>.492</td>
<td>.040</td>
</tr>
<tr>
<td>LCCr</td>
<td>1.12 ± 0.05</td>
<td>1.10 ± 0.06</td>
<td>-1.46</td>
<td>.237</td>
<td>.145</td>
</tr>
<tr>
<td>LSCr</td>
<td>2.73 ± 1.25</td>
<td>2.84 ± 1.34</td>
<td>-0.65</td>
<td>.106</td>
<td>.515</td>
</tr>
</tbody>
</table>

Note: significant differences according to our adjusted alpha (p < .0167) are shown in red, while those that are between our adjusted alpha and the traditional significance level (.0167 < p < .05) are shown in green.
Figure 5. Boxplots showing distribution of participant median values of TRC, Edges, LCC in REM and N2 sleep. Significant differences (* .0167 < p < .05; ** .0167 < p) from Wilcoxon sign-rank tests are shown.
**Testing for time of night effect**

For our third hypothesis, we aimed to investigate whether graph connectedness and random-likeness changed across the night. To test this, we investigated whether TRC and variables of graph structure (Edges, LCC, LSC, LCCr, LSCr) could predict the order of awakening in which dream reports were obtained. We first entered sleep stage in as a variable for model comparison, since we are interested in whether changes across the night are observed independent of any residual differences that exist between the sleep stages. As a result, variables of interest (Edges, LCC, LSC, TRC, LCCr, LSCr) were entered individually to a model containing sleep stage in order to investigate whether their addition improved the overall fit of the model. From the resultant models, none of the variables were found to significantly improve the overall fit. This is also evident in an observed increase in AIC values in all resultant models (except for Edges), which indicates that the overall fit of the model was not improved by the addition of the respective variables. This suggests that graph structure and TRC was not related to the order of awakenings. Given these results, no time of night effect was found and our third hypothesis was not corroborated.
Table 4
Output for Cumulative Link Models Estimating Relationship to the Order of Awakenings

<table>
<thead>
<tr>
<th>All Reports</th>
<th>AIC</th>
<th>Pseudo R²</th>
<th>Pseudo R² Change</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Model</td>
<td>504.43</td>
<td>0.00</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Sleep Stage</td>
<td>501.71</td>
<td>.038</td>
<td>.038</td>
<td>-2.36</td>
<td>.030</td>
</tr>
<tr>
<td>Sleep Stage + Edges</td>
<td>500.64</td>
<td>.062</td>
<td>.025</td>
<td>-1.54</td>
<td>.080</td>
</tr>
<tr>
<td>Sleep Stage + LCC</td>
<td>503.35</td>
<td>.041</td>
<td>.003</td>
<td>-0.18</td>
<td>.552</td>
</tr>
<tr>
<td>Sleep Stage + LSC</td>
<td>501.95</td>
<td>.052</td>
<td>.014</td>
<td>-0.88</td>
<td>.185</td>
</tr>
<tr>
<td>Sleep Stage + TRC</td>
<td>503.01</td>
<td>.03</td>
<td>.005</td>
<td>-0.35</td>
<td>.402</td>
</tr>
<tr>
<td>Sleep Stage + LCCr</td>
<td>502.99</td>
<td>.043</td>
<td>.006</td>
<td>-0.36</td>
<td>.397</td>
</tr>
<tr>
<td>Sleep Stage + LSCr</td>
<td>503.42</td>
<td>.040</td>
<td>.002</td>
<td>-0.14</td>
<td>.593</td>
</tr>
</tbody>
</table>

*Note: Pseudo R² change values are calculated in comparison to a model containing *sleep stage*, while Pseudo R² are calculated in relation to the null model. Values that are between our adjusted alpha and the traditional significance level (.0167 < p < .05) are shown in green.
Distinguishing sleep stage based on graph structure and TRC

For our fourth hypothesis, in order to test how graph structure compares to TRC in sleep stage discernment, we constructed generalised linear models with a binomial (REM/N2) outcome in order to examine whether aspects of graph structure can significantly distinguish between reports obtained from REM and N2 sleep and how they compare to TRC. While such a test is theoretically similar to the Wilcoxon sign-rank tests, which aimed to detect differences in REM and N2, the generalised linear model possesses the advantage of utilising the full sample of dream reports (n = 125) and can evaluate how certain predictor variables relate to one another in modelling a certain outcome.

Testing individual variables

To test the ability of the individual variables, measures of graph structure (Edges, LCC, LSC, LCCr, LSCr) and TRC were entered individually as predictor variables. The analyses found that the addition of LCC and TRC significantly improved the model in predicting differences in REM and N2 (see Table 5, Figure 6). The differences found after adding Edges, LSC, LCCr, and LSCr did not significantly improve the null model and thus suggested that they could not significantly predict sleep stage. Thus, mirroring the significant differences found in our Wilcoxon-sign rank tests, we found that the variables TRC and LCC were the best performing variables in detecting differences amongst REM and N2 reports. Given that some (LCC and TRC) but not all variables could significantly discern sleep stage, our fourth hypothesis was only partially corroborated.

Testing for complementary measures

As a follow up to the significant findings for LCC and TRC, we next investigated whether LCC and TRC can act as complementary measures to one another in the discernment of
sleep stage. In this regard, we tested whether the addition of LCC to a model containing TRC would significantly improve the fit of the model in predicting differences in sleep stage. To do this, we compared the log-likelihood ratios of a model containing TRC with and without the addition of LCC as a fixed effect. The model containing both TRC and LCC was found to be a significantly better at predicting sleep stage than TRC alone (see Table 5). We performed the same analysis, this time seeing whether TRC could add significantly to a model containing LCC. Once again the difference between the models was significant, suggesting that both TRC and LCC are complementary measures in sleep stage discernment.
Table 5
Output for Logistic Regression Predicting Sleep Stage Through Graph Connectivity and Report Length

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>AIC</th>
<th>Pseudo R²</th>
<th>Pseudo R² Change</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Model</td>
<td>168.5</td>
<td>0.00</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Edges</td>
<td>168.2</td>
<td>0.25</td>
<td>.025</td>
<td>-1.151</td>
<td>.129</td>
</tr>
<tr>
<td>LCC</td>
<td>162.1</td>
<td>.088</td>
<td>.088</td>
<td>-4.178</td>
<td>.003</td>
</tr>
<tr>
<td>LSC</td>
<td>169.7</td>
<td>.008</td>
<td>.008</td>
<td>-0.379</td>
<td>.384</td>
</tr>
<tr>
<td>LCCr</td>
<td>168.27</td>
<td>.002</td>
<td>.002</td>
<td>-0.100</td>
<td>.654</td>
</tr>
<tr>
<td>LSCr</td>
<td>167.68</td>
<td>.009</td>
<td>.009</td>
<td>-0.395</td>
<td>.374</td>
</tr>
<tr>
<td>TRC</td>
<td>159.7</td>
<td>.113</td>
<td>.113</td>
<td>-5.397</td>
<td>.001</td>
</tr>
<tr>
<td>TRC + LCC</td>
<td>155.7</td>
<td>.179</td>
<td>.066</td>
<td>-2.980</td>
<td>.015</td>
</tr>
<tr>
<td>LCC + TRC</td>
<td>155.7</td>
<td>.179</td>
<td>.091</td>
<td>-4.200</td>
<td>.003</td>
</tr>
</tbody>
</table>

Note: Significance testing and Change in Pseudo R² are calculated in comparison to the Null Model for the first set of individual measures, and calculated in comparison to a model containing either TRC or LCC in the composite analyses. Values that reach statistical significance according to our adjusted alpha (α < .0167) are shown in red.
Figure 6. Graphs showing predicted probabilities of sleep stage based on TRC, Edges, LCC and LSC. TRC and LCC were found to be statistically significant.
Testing the relationship to dream complexity

Individual variables

For our fifth hypothesis, we aimed to evaluate whether TRC and measures of graph structure are related to external ratings of dream complexity (the PIRS). To investigate this, as before, significance values were calculated by comparing the log-likelihood ratios of models with and without the desired fixed effect in question. The null model adopted for comparison contained the fixed effect of sleep stage, since we are interested whether the explanatory variables can significantly improve the fit of the model over and above differences in complexity present between the sleep phases.

The results for the analysis can be found in Table 6. The addition of Edges, LCC, TRC and LCCr to a model containing sleep stage were all found to significantly improve the fit of the model in predicting PIRS scores, while LSC showed a significant trend in the same direction; LSCr was not found to be statistically significant. In terms of the direction of this relationship, the results indicate that report length and graph connectedness increases, while graph random-likeness decreases in relation to increased ratings of dream complexity. The effect sizes of graph structure measures, as estimated by a change in Nagelkerke's pseudo-$R^2$, were found to be of a small to medium size; the effect size for the addition of TRC was large. In order to test whether the slope of effect in predicting dream complexity was different in REM or N2, we tested for the presence of an interaction effect between sleep stage and the fixed effects in the respective models (TRC, Edges, LCC, LSC, LCCr, LSCr). The addition of the interaction effect to these models did not significantly improve the overall fit (TRC: AIC = 349.52, Pseudo $R^2$ Change = .020, $\chi^2 = -1.17$, p = .127; Edges: AIC = 426.39, Pseudo $R^2$ Change = .011, $\chi^2 = -0.640$, p = .258; LCC: AIC = 420.65, Pseudo $R^2$ Change = .002, $\chi^2 = -0.10$, p = .651; LSC: AIC = 429.99, $\chi^2 = -1.17$, p = .127; LCCr: AIC = 420.65, Pseudo $R^2$ Change = .002, $\chi^2 = -0.10$, p = .651; LSCr: AIC = 429.99, $\chi^2 = -1.17$, p = .127).
Pseudo $R^2$ Change = .003 $\chi^2 = -0.17$, p = .559). As a result, we can assume that the trends for REM and N2 groups were not significantly different from one another in their prediction of dream complexity. Given that TRC aspects of graph connectedness and its random-like nature could significantly predict ratings of dream complexity, our fifth hypothesis was corroborated.

**Testing for complementary measures**

Given the significant relationships found, we next sought to investigate whether attributes of graph structure that were found to be previously significant could act as complementary measures to TRC in explaining dream complexity. To do so, we compared the log-likelihood ratios of a model containing TRC and the individual connectedness measures to a model only containing TRC. We found that the addition of LCC significantly improved the fit of the model; no such effect was found in either Edges, or LCCr. As a result, only LCC acted as a complementary measure to TRC in explaining differences in dream report complexity.
Table 6

*Output for Cumulative Link Models Estimating Relationship Between Graph Connectedness and Dream Complexity (PIRS)*

<table>
<thead>
<tr>
<th></th>
<th>All Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2.1 Individual Measures</strong></td>
<td></td>
</tr>
<tr>
<td>Null Model</td>
<td>453.15</td>
</tr>
<tr>
<td>Sleep Stage</td>
<td>430.76</td>
</tr>
<tr>
<td>Sleep Stage + Edges</td>
<td>425.67</td>
</tr>
<tr>
<td>Sleep Stage + LCC</td>
<td>420.65</td>
</tr>
<tr>
<td>Sleep Stage + LSC</td>
<td>428.33</td>
</tr>
<tr>
<td>Sleep Stage + TRC</td>
<td>349.85</td>
</tr>
<tr>
<td>Sleep Stage + LCCr</td>
<td>421.16</td>
</tr>
<tr>
<td>Sleep Stage + LSCr</td>
<td>432.69</td>
</tr>
<tr>
<td><strong>2.2 Composite Models:</strong></td>
<td></td>
</tr>
<tr>
<td>Sleep Stage + TRC + LCC</td>
<td>344.44</td>
</tr>
<tr>
<td>Sleep Stage + TRC + Edges</td>
<td>353.45</td>
</tr>
<tr>
<td>Sleep Stage + TRC + LCCr</td>
<td>353.34</td>
</tr>
</tbody>
</table>

Note: Values that reach statistical significance (adjusted $\alpha < .0167$) are highlighted in red; significance values that fall in the range of $(.0167 < p < 0.05)$, are highlighted in green. Values of Pseudo $R^2$ Change are calculated in comparison to the sleep stage model for individual measures and in comparison to the model containing TRC for the composite ones. Values of Pseudo $R^2$ are calculated in comparison to the null model.
Dependence on dream complexity in predicting sleep stage

Given that our results indicate that LCC and TRC can predict differences in sleep stage (REM vs. N2) and ratings of dream complexity, we added a supplementary hypothesis to our study that sought to investigate whether the ability of LCC and TRC in to discern sleep stage is independent of differences in ratings of dream complexity. In this sense we are trying to evaluate statistically whether the differences previously found in REM and N2 are independent of underlying differences in dream complexity. In order to investigate this, we returned to our generalised linear model and partialled out inherent differences in dream complexity through testing whether the overall fit of a model containing PIRS would be significantly improved by adding LCC or TRC to the model. By comparing the log-likelihood ratios of the respective models, we found that the addition of either LCC (AIC = 156.0, Pseudo R² Change = .028, \( \chi^2 = -1.162, p = .127 \)), TRC (AIC = 158.3, Pseudo R² Change = <.001, \( \chi^2 = -0.001, p = .959 \)) or both LCC and TRC (AIC = 158.0, Pseudo R² Change = 0.028, \( \chi^2 = -1.179, p = .307 \)) did not significantly improve the fit of the model in predicting sleep stage. This indicates that the ability of LCC and TRC to distinguish sleep stage in this sample is dependent on underlying differences in dream complexity.

Model checking

No issues were found in the variance inflation factor (all VIF values < 4) when attributes were entered in collectively in the composite models, indicating no issues in collinearity. An important assumption to be checked in cumulative link models is the proportional odds assumption, which states that the effect of the predictor variables in one’s model is constant across all levels of the outcome variable. In order to check for this assumption, we used the nominal test functions within the ordinal package (Christensen, 2015). The nominal test was
found to be non-significant for all of our models, except for the variable sleep stage (p < .05) indicating that there may be issues with an assumption violation in the model relating to this variable. In order to check whether such a violation may have affected our findings, we performed the same analyses employing a partial proportional odds model (‘clm2’ function), specifying the variable sleep stage as a nominal effect in order to control for any previous assumption violations related to this variable. The significance of the resultant models for Edges, LCC, TRC, and TRC + LCC remained significant, while LSC (p = 0.08) was found to be not-significant. Given that LSC did not originally reach statistical significance according to our adjusted alpha (α < .0167), the respective changes found in the partial proportional odds models did not impact our overall findings in anyway. As a result, we can be confident that the findings presented here are not erroneously skewed due to a violation of the proportional odds assumption.
Discussion

This research aimed to investigate the differences in the structural organisation of REM and non-REM dream reports through speech graph analysis. This is the first study to demonstrate that REM reports possess a larger connectedness as compared to N2 reports, a difference that reflects words recurring over a longer range in REM reports. It also indicates that graph structure, both in terms of connectedness and its random-likeness, is informative of the dream complexity described in the report, where more complex dreams are associated with larger connectedness and less random-like graph structures. Finally, the results demonstrate that aspects of graph connectedness (specifically LCC) can act as a complementary measure to TRC in predicting differences in REM and non-REM dream reports and overall ratings of dream complexity. Collectively, these results complement current literature reporting differences in REM and non-REM dreaming; point to SGA as a promising automated measure for future use in dream research. In the following sections, we discuss the present findings in relation to past literature reporting differences in REM and non-REM dreaming; explore possible explanations for the observed differences graph connectedness; and propose that SGA possesses particular value as a method for future use in dream research.

REM Reports Are Longer and Have Larger Connectedness Compared to N2

The results found in the present study replicate a number of findings in previous studies pointing to differences in REM and non-REM dreaming. Firstly, we found that dream recall is higher in REM than during N2, supporting findings that dreaming is more frequent in REM (Nielsen, 2000). Secondly, we found that qualitatively, REM dreams were more often rated as being part of an ongoing narrative, while non-REM dreams often involved non-visual, conceptual recall. This is line with previous studies showing that REM dreams are more visually
hallucinatory (Fosse et al., 2004) and story-like (Nielsen et al., 2001), while non-REM dreams are often thought-like (Fosse et al., 2004) and conceptual (Foulkes & Rechtschaffen, 1964). Finally, in our sample, REM reports were typically longer than N2 ones (i.e. higher TRC), supporting previous studies showing that one of the most robust differences found between these two groups relates to longer report length (Antrobus, 1983; Casagrande et al., 1963; Goodenough et al., 1965; Fosse et al., 2001; Foulkes & Rechtschaffen, 1964; Oudiette et al., 2012; Stickgold et al., 1994; Wamsley et al., 2007).

By using the sliding window method in order to control for these report length differences, we aimed to investigate whether intrinsic structural differences are found between these two groups, independent of residual differences report length. The results found that REM reports had larger connectedness compared to N2 in term of LCC, while Edges and LSC showed a trend towards significance in this same direction. On the other hand, when comparing dream reports to reports that were randomly shuffled 1000 times, we didn’t find any differences in REM and non-REM reports in their random-like nature. This suggests that, on average, words contained within REM reports tend recur with a longer range compared to those of N2, forming longer loops and far-reaching connections. However, they suggest that these structural differences are not accompanied by differences in the way that they approximate to random speech, such as that which is found in people suffering from schizophrenia (Mota et al., 2017). While these results are in accordance with findings from previous studies, we were not able to replicate findings from studies that have reported a time of night effect (Antrobus et al., 1995; Wamsley et al., 2007). While graph connectedness did not change as a factor of the order of awakening, no such effect was found TRC either. Since TRC has been previously shown to increase across the course of the night (Antrobus et al., 1995; Wamsley et al., 2007), such
findings are inconsistent with the previous literature. We interpret this to be a reflection of the use of a measure (i.e. order of awakenings) that is too broad and imprecise to be able to detect such an effect, since the number and time of the awakenings were not held constant across the experimental nights. Indeed, by using the order of awakening, Oudiette et al. (2012) also found no effect on measurements of dreaming across the night. Thus, as opposed to being a finding that genuinely contradicts those of previous studies, we feel that a more precise measurement of time (e.g. clock-time/time since sleep onset) is needed in order to clarify how graph connectedness may/may not evolve in relation to the time of night.

As a whole, these results collectively suggest that dream reports are less frequent in N2, and when they are present, they are typically shorter, more thought-like and have smaller connectedness compared to their REM report counterparts. Given that many differences in REM and non-REM reports are highly diminished or even disappear after controlling for report length (Antrobus, 1983), these findings also possess value in supplementing a small group of studies that have found differences between these sleep stages over and above residual differences in report length (Casagrande et al., 1996; Speth et al., 2013; Waterman et al., 1993). Further research may investigate the time of night effect in order to clarify whether graph connectedness increases across the night in a similar fashion to TRC.

**Graph Connectedness in Relation to Dream Reports Across the Sleep Cycle**

Previous studies have found that SGA in dream reports can be particularly informative of the thought disturbances that underlie psychosis (Mota et al., 2012; Mota et al., 2014, Mota et al. 2017). Such findings naturally prompt comparisons to the long-held phenomenological comparisons between dreaming as a model for psychosis (Mota et al., 2014). One of the hallmark differences between REM and non-REM dreaming is due to its bizarre, hallucinatory
nature (Fosse et al., 2004). By extension, if we are to assume that REM dreaming more closely approximates this model in terms of its bizarreness and hallucinatory nature, one may come to speculate that graphs obtained from REM reports would more closely approximate those of people with schizophrenia (i.e. would be less connected). However, such an interpretation seems inconsistent with the present findings, since it runs directly against the direction of our results, where REM graphs had on average larger connectedness compared to N2 graphs, and not the other way round. If we were to apply this framework to our sample, it would suggest that N2 dream reports approximate the reports of those with psychosis more than REM, which seems improbable according to its phenomenological nature. Thus, while the phenomenological aspects of dreaming may approximate the experiences of people with psychosis, the differences in connectedness of dream reports across the sleep cycle in healthy young adults is not reflected in this regard.

We believe a more suitable approach to the present data would be to interpret the observed differences in graph connectedness in terms of variations in the cognitive ability of participants to retrieve and organise their dream experiences. This is in accordance with findings that graph connectedness tends to increase in healthy cognitive development in children (Mota et al., 2016), and declines in age-related dementias (Bertola et al., 2014) and forms of psychopathology (Mota et al., 2012, Mota et al., 2014, Mota et al., 2017) where cognitive impairment is commonly observed. For the present study, we postulate that the observed changes of graph connectedness in dream reports across the sleep cycle may be conceivably affected by two main factors.

The first factor to be considered is related to sleep inertia and the immediate effects of the sleep/wake transition, whereby memory and attention processes may be impaired. Since dream
report collection was conducted shortly after controlled awakenings, it is possible that such a transition may influence graph connectedness by impairing one’s ability to organise one’s thoughts and memories in and is reflected in smaller graph connectedness. Since sleep inertia is more marked in N2 compared to REM (Cavallero & Versace, 2003), one can imagine that sleep inertia may exert a more negative impact on the ability to organise one’s thoughts in N2, leading to an apparent decrease in report connectedness compared to REM.

The second factor to be considered is related to the nature of the dream experience itself. Since the quality of dreaming may vary considerably both within and between sleep states, it is possible that the ability to organise one’s experience into a report may be influenced by the underlying complexity of the dream experience that is to be described. In this sense, dream experiences that are coherent and story-like may be more easily organised into a report with larger connectedness, while dream experiences that are fragmented and isolated are relatively more difficult mentally organise and thus are structurally smaller in connectedness.

While the role of sleep inertia cannot be completely ruled out by the present study, the results we obtained tend to favour the second interpretation for a number of reasons. Firstly, once we partialled out differences in dream complexity, as rated by the PIRS, the ability of LCC to distinguish REM and N2 dream reports was not statistically significant (see 3.2.3). This suggests that the ability of LCC to discern REM and N2 dream reports is dependent on the underlying differences in dream complexity found between these two sleep stages. Secondly, by using a model containing sleep stage as a statistical comparison, we could investigate whether graph connectedness could significantly predict PIRS over and above any differences in sleep stage (i.e. when any differences related to the sleep stage are partialled out). In our analysis, several variables of graph connectedness were found to significantly improve the sleep stage model in
predicting ratings of dream complexity, indicating that graph connectedness is related to dream complexity, independent of any inherent differences in graph connectedness from the REM and N2 sleep stages. Furthermore, no significant interaction effect was found between the graph attributes and sleep stage as a variable, indicating that the modelled relationship between TRC and graph connectedness with PIRS was comparable for both the REM and N2 groups of dream reports. This suggests that within-group differences in graph connectedness of REM and N2 dream reports are comparably related to the overall ratings of dream complexity. As a result, unless there is an intrinsic connection between the intensity of sleep inertia experienced following an awakening and the overall rated complexity of the dream experience, the interpretation that sleep inertia can explain the present findings seems to be insufficient.

As an alternative, given the results described above, a more empirically-grounded explanation of the present findings would be to interpret graph connectedness as a reflection of underlying differences in dream report complexity. In this regard, dreams that are more complex and involve coherent, story-like experiences are more easily organised into a larger connected and non-random report structure, while dreams that are isolated and incoherent are more difficult to mentally organise and are reflected as smaller connectedness and being more random-like. Such an explanation can also explain the observed REM and N2 differences in graph connectedness, since complex story-like dreams, are more common in, but not exclusive to, REM sleep. This hypothesis may be corroborated/falsified in future research through investigating the relationship between the narrative/story-like complexity of dreams and their graph connectedness in different samples. Since the narrative complexity of dream reports tends to persists even after a period of time has elapsed (Cipolli et al., 1998), one may uncouple the effects of sleep inertia from dream complexity through analysing and comparing the story-
likeness and structural connectedness of reports obtained immediately after an awakening to another set of reports that describe the same dream experiences from during the night, only at a delayed time, perhaps the following morning, where any residual cognitive effects of the sleep/wake transition can be considered to be greatly diminished, if not completely ruled out. Should connectedness and story-likeness remain the same across both conditions, it would corroborate the findings presented here, while if differences are found whereby connectedness increases in the morning report, despite the story-like complexity remaining constant, it would favour an influence of sleep inertia and go against the interpretation present here. Clearly, since these explanations are not mutually exclusive, graph connectedness may be affected by a combination of these factors, as well as other factors that may have not been considered here.

**SGA as a Method for Dream Research**

By utilising hierarchical model construction in discerning sleep stage (REM vs. N2) and levels of dream complexity (as measured by the PIRS), we were able to test how graph connectedness compared to TRC in modelling these variables of interest and whether it could act as a complementary measure in this regard. We found that not only could LCC predict differences in sleep stage and ratings of dream complexity, but also that it could significantly improve a model containing TRC in this prediction. Given that TRC is considered to be one the most widely used measures to distinguish REM and non-REM reports, this finding is of particular important since it suggests that LCC can act as a complementary measure to TRC in discerning the sleep stage and relevant level of dream complexity of a report. While Edges, LSC and the random-likeness of LCC (LCCr) could not significantly discern REM and non-REM dreams or significantly improve models containing TRC, they still showed promise in predicting differences in dream complexity.
As a whole, these findings point to SGA as a promising tool for dream research. The automatized nature of SGA means that it is fast, economical and avoids the biases and problems of reliability inherent to methods that involve human rating systems. Given that in the analysis of dream reports, the most common methods include the use of human judges who are trained to rate dream reports according to a given scale (e.g. Hall & Van de Castle, 1966), SGA offers an unbiased alternative to these traditional measures since does not introduce problems related to human bias. Furthermore, the fact that it is automatic and freely available means that it can be applied to large corpora of dream reports, that may otherwise be too time-consuming and/or expensive to apply traditional, human-based rating systems. The advent of the Dream Bank (Domhoff & Schneider, 2008), which now holds more than 20 000 dream reports means that such benefits are already applicable.

While SGA is very much in its infancy in its application to dreaming, this study represents a first step in this direction and future studies can help evaluate SGA’s potential use in different samples of dream reports, perhaps in longer reports, which would allow one to explore a larger sized windows (e.g. 50 words). Further investigations can examine SGA’s scope of use by applying it to different populations, settings (e.g. dream collected in the laboratory vs. at home) and report formats (written vs. verbally reported). Research conducted in this regard can help further our understanding of graph structure as an informative measure of nocturnal mental activity and its potential to complement report length in dream report analysis.

**Limitations and Perspectives for the Future**

In light of the present findings, a number of limitations need to be considered that could have improved our study and the interpretation of our findings.

Firstly, as has been mentioned before, it is unclear how sleep inertia may have affected
the graph connectedness results. While we have shown that statistically such an influence is unlikely to be able to explain our findings, such an influence cannot be ruled out. This can be clarified by future studies by comparing verbal dream reports from controlled awakenings obtained immediately after the awakening and at delayed intervals, where such effects can be assumed to have worn off. Since attentional and memory processes are at their peak during the day, waking reports may also be particularly helpful as a control comparison in this regard.

Secondly, this study could have been improved by including a more precise measure in estimating the time of night effect in relation to graph connectedness and TRC. While the order of awakening is a theoretically related to the time of night, the number of awakenings was not fixed across the experimental nights and awakenings were conducted at different times. The use of this measurement was due to the unavailability of the specific time of awakenings. A number of previous studies have reported an overall increase in dream complexity across the night in both REM and N2 dream reports (Antrobus et al., Wamsley et al., 2007). By including the waking time of participants, future studies investigate whether graph connectedness is associated with the same time of night increases as is observed in report length and ultimately improve the overall fit of the models presented here.

Thirdly, our participant median TRC estimates in REM (49) and N2 (35) are closer to one another compared to those cited in previous studies (e.g. Oudiette et al., 2012, REM - 40, N2 - 13; Stickgold et al., 1994, REM - 148, N2 - 21). Thus, it is possible that TRC’s potential as a measure to predict differences in sleep stage may be diminished here due to to inherent characteristics of the sample. This can be remedied by applying SGA to other representative samples, where it’s ability to distinguish REM from non-REM may be meaningfully compared to TRC and other traditional measures of dreaming.
Fourthly, unlike other studies using graph analysis where the dialogue between researcher and experimental subject is time-limited (e.g. Mota et al., 2014), the present study used previously collected dream reports and thus had no comparable control. However, the protocol for the dialogue used here is well established and was strictly adhered to in collecting the dream reports. Furthermore, since investigating speech structure was not a part of the aims and hypotheses of the original study, any potential experimenter bias related to influencing speech structure can be ruled out. Nonetheless, using a time-limited control may be a useful consideration for future studies.

Finally, we should consider that while we have reported differences in REM and non-REM reports, the scope of our findings is restricted to N2 reports. Given the inherent differences in the types of mentation experienced across the non-REM stages (see Nielsen, 2000, for a review), we cannot generalise our findings here to the other N1 and N3 sleep stages. Future studies incorporating N1 and N3 reports, as well as waking mentation reports for comparison can enhance our understanding of changes in graph connectedness across the sleep/wake cycle.
Conclusion

In the present study, we have shown, via SGA, that the word-to-word structural organisation of dream reports is informative about the sleep stage in which it was obtained and the overall complexity of the dream report, even when differences in report length are partialled out. As a whole, the results replicate previous findings showing that dreaming in N2, as compared to REM, is less frequently recalled and, when present, is shorter, less intense and more thought-like and conceptual; they also supplement previous research by extending these findings to show that N2 reports display smaller connectedness (i.e. words recur over a shorter range), compared to their REM report counterparts. Although a time of night effect has been found in previous literature, we were not able to replicate the finding here, probably owing to the imprecise nature of the measurement we used. While the effects of sleep inertia cannot be ruled out, the observed differences in graph structure appear to be a reflection of underlying differences in the dream complexity, where coherent, story-like dream experiences (more commonly found in REM), are more likely to be organised with larger connectedness and a less random-like report structure. Such findings represents a significant step towards characterising the evolution of the structure of mentation reports across the various phases of the sleep cycle. They also point to SGA as a promising automated measure for sleep research due to its relationship to dream complexity and its ability complement report length in the analysis of REM and non-REM dream reports. Further research can replicate and extend these findings through clarifying the effects of sleep inertia on graph connectedness and evaluating the evolution of graph structure according to the time of night effect. Such investigations can enhance our knowledge of dreaming and its various manifestations throughout the night, while providing further evidence for the application of automated graph-based methods in dream research.
References


Cavallero, C., & Versace, F. (2003). Stage at awakening, sleep inertia and performance. *Sleep*


Lawrence Erlbaum Associates.


Rosenlicht, N., Primich, C., McQuaid, J. R., Maloney, T., & Feinberg, I. (2017). Dreamlike events are correlated with the length of sleep mentation reports. *Archives italiennes de biologie, 155*(1-2), 64.


collection of dreams. *Journal of sleep research*, 2(1), 8-12.


Supplementary Material
Appendix A: An Illustration of the Sliding Window Method
Ok, ah, I can remember. It was sort of outside, um... I think next to ah, a park, somewhere... ah, yeah, and ah, we were just sitting there, talking, and... um, there was a point when I was like the only person who had ah, things actually on his head and his body, yeah, I think so.

Ah, like ah, they were cords, like the ones on, like these ones.

Yes electrodes.

---

Figure A1. An illustration of the sliding window method (15 word window, 10 word overlap)

Figure A2. Formulas for Calculating Sliding Window Scores

\[
\text{Edges} = \frac{\sum \text{Edges Scores in All Windows}}{\text{No. of Windows}}
\]

\[
\text{LCC} = \frac{\sum \text{LCC Scores in All Windows}}{\text{No. of Windows}}
\]

\[
\text{LSC} = \frac{\sum \text{LSC Scores in All Windows}}{\text{No. of Windows}}
\]
Appendix B: An Illustration of Randomised Word Graphs
Figure B1. An illustration of graph randomisation

Original Report (Before Word Shuffling)

hmm... I just remember my grandmother. Um... and that was it; I don’t really remember what she was doing or anything, I just remember she was there.

Random Report (After Word Shuffling)

hmm... doing just there my what. Um... and she was just; I don’t I remember grandmother that was I or anything, really it remember she was remember.

Figure B2. Formulas for Calculating Random Attribute Scores

\[
\text{LCCr} = \frac{\text{LCC(Original graph)}}{\text{mean(LCC(1000 random graphs))}}
\]

\[
\text{LSCR} = \frac{\text{LSC(Original graph)}}{\text{mean(LSC (1000 random graphs))}}
\]
Appendix C: Confirmation of Ethics Clearance
Ms Danyal Wainstein
Department of Psychology
University of Cape Town
Rondebosch 7701

15 March 2017

Dear Ms Wainstein

I am pleased to confirm that ethical clearance was given by an Ethics Review Committee of the Faculty of Humanities for your study, "Non-REM dreaming in relation to the cyclic alternating pattern (CAP): An exploratory study". The ethical clearance commenced in 2011, and will remain active until the study is completed.

I wish you all the best for your study.

Yours sincerely,

Lauren Wild (PhD)
Associate Professor
Chair: Ethics Review Committee
Appendix D: Supplementary Statistics
Table E1

*Shapiro-Wilk normality tests in report length and graph attribute measures*

<table>
<thead>
<tr>
<th></th>
<th>W-Statistic</th>
<th>Significance Test (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRC</td>
<td>0.86</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Edges</td>
<td>0.77</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LCC</td>
<td>0.95</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LSC</td>
<td>0.92</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LCCr</td>
<td>0.94</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LSCr</td>
<td>0.91</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Note:* Significant test results shown in red indicate that data is not normally distributed.

Table E2

*Pearson correlation between predictor variables*

<table>
<thead>
<tr>
<th></th>
<th>TRC</th>
<th>Edges</th>
<th>LCC</th>
<th>LSC</th>
<th>LCCr</th>
<th>LSCr</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRC</td>
<td>1.00</td>
<td>0.60</td>
<td>0.31</td>
<td>0.49</td>
<td>-0.94</td>
<td>-0.77</td>
</tr>
<tr>
<td>Edges</td>
<td>1.00</td>
<td>0.60</td>
<td>0.94</td>
<td>-0.66</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>LCC</td>
<td>1.00</td>
<td>0.49</td>
<td>-0.14</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSC</td>
<td>1.00</td>
<td>-0.60</td>
<td>-0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCCr</td>
<td>1.00</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSCr</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* Values are indicative of Pearson's r correlation coefficient.
Figure E1. Correlation plot showing spearman correlations between predictor variables. Blue colours indicate a positive relationship between the two, while red colours indicate a negative relationship.
Figure E2. Boxplot showing inter-participant differences in report length and graph attributes.
Figure E3. Boxplot showing between-night differences in report length and graph attributes.