Social Networks 74 (2023) 166–181

Contents lists available at ScienceDirect

Social Networks

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

What is the Point of Change? Change Point Detection in Relational Event Models

Mahdi Shafiee Kamalabad [a,e,*], Roger Leenders [c,d], Joris Mulder [b,d]

[a] Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands
[b] Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands
[c] Department of Organization Studies, Tilburg University, Tilburg, The Netherlands
[d] Jheronimus Academy of Data Science, 's-Hertogenbosch, The Netherlands
[e] Centre for Complex Systems Studies, Utrecht University, The Netherlands

ARTICLE INFO

Keywords:
Social network analysis
Change point detection
Relational event model
Bayes factor
Communication
Critical situation
Critical incident analysis
Resilience engineering
Graceful extensibility

ABSTRACT

This paper presents an extension to the relational event model with change points (REM-CP) to study abrupt changes to social interaction behavior in temporal networks. A change point detection algorithm is proposed for exploring when and which network effects abruptly change, and a confirmatory approach to test the presence of a change point at a given moment. The effectiveness of the methodology was assessed with numerical simulations and NASA's Apollo 13 mission data. The latter revealed dynamic communication behavior and identified time zones where most change points occurred, including around the time of the famous quote "Houston, we've had a problem."

1. Introduction

It was expected to be a fairly routine mission. By the time the Apollo 13 mission was scheduled, it would already be the seventh crewed exploration mission in the Apollo program. As a result, NASA had accumulated lots of experience on how to run these.

Apollo was a NASA program in which for the first time people had left Earth and had landed on the Moon. For the most part, the Apollo 13 mission would do what others had done before them already and for which ample experience and appropriate procedures had been accumulated. The mission's innovation would be that the astronauts were going to conduct a geological survey of the moon's surface, but anything before or after that was just going to be a repetition of what had been done before. At least, that is what everybody thought.

The craft was launched on April 11, 1970, from Kennedy Space Center and the first two days were indeed uneventful and standard. Then, a routine stir of one of the oxygen tanks caused an explosion that vented the oxygen of both of the Service Module's oxygen tanks into space. As a result, life support for the astronauts, electric power, and propulsion were gone. Routine was out the window too. Apollo 13 was effectively shipwrecked in space. Command module pilot Jack Swigert was the first to report something was wrong: "Houston, we've had a problem".

What followed was an intense search for a way to bring the astronauts back home. One of the challenges was figuring out how to connect square-shaped connections to round connections using plastic bags, cardboard, duct tape, and even socks. The stuff that Hollywood movies are made of. While the communication among the crew members and between the crew and mission control had been quite routine and stable until the accident, the communication patterns were quite different from those of the first two days but crucial to solve this unexpected and unprecedented situation van den Oever and Schraagen (2021). Understanding how teams communicate to solve critical situations such as these can help understand teamwork and train other teams in bringing home their metaphorical spaceships. To give the reader an impression of this relational event sequence, Table 1 presents the first 6 communication messages between the actors in the network.

The relational event model (REM) (Butts (2008)) has become a well-established approach to understanding communication structures based on observed social interactions in real-time. The REM parameterizes the interaction rates between actors as a function of endogenous and exogenous statistics, given the event history. The endogenous statistics summarize the information in the system (network) up until a given time point, say t, while the exogenous statistics contain external, typically static, information (such as actors' attributes). The

* Correspondence to: Padualaan 14, 3584 CH, Utrecht, The Netherlands.
E-mail address: m.shafieekamalabad@uu.nl (M. Shafiee Kamalabad).

https://doi.org/10.1016/j.socnet.2023.03.004

Available online 4 April 2023

Table 1
First six relational interactions of Apollo 13 data. The times are formatted as hh:mm:ss. We shortened the first message due to the lack of space.

Time (hh:mm:ss)	Sender	Receiver	Message
54:46:28	CMP	CAPCOM	Okay, Okay, Houston, we have transferred to LM power; …
54:46:40	CAPCOM	CMP	Roger, Jack. Transfer to LM power. Thank you.
54:46:47	LMP	CAPCOM	And the docking-tunnel index, Jack, was minus 2 degrees.
54:46:54	CAPCOM	LMP	Say again, Fred, you are coming in with a lot of background noise.
54:47:00	LMP	CAPCOM	Okay. The docking-tunnel index mark was minus 2 degrees.
54:47:05	CAPCOM	LMP	Roger. Minus 2 degrees.

corresponding effects explain how much these statistics affect social interaction behavior in the network.

The REM can be used to study fine-grained temporal social network dynamics and estimate how specific drivers (e.g., inertia, reciprocity, two-paths, and status hierarchy) appear to shape which interactions occur and when they occur. The conventional REM assumes that the effects of the endogenous and exogenous statistics are constant over time. On the other hand, as shown from the Apollo 13 example above, it is plausible that in some situations, the strengths of some of the effects (and, hence, the social dynamics themselves) change at certain points in time. The conventional REM does not identify the number and locations of the time points where the effects strengths change abruptly. That is, given the past event history, the REM does not pinpoint at which points in time the communication dynamics change qualitatively. The ability to observe such change points (CPs) may improve our understanding of instantaneous changes in communication behavior over time.

Change points (CPs) are time points where the strengths of effects in a model suddenly and strongly change, Shafiee Kamalabad and Grzegorczyk (2020). To the best of our knowledge, relatively little work has been carried out on developing a change point (CP) detection method to detect the number, locations, and size of CPs in REMs. However, the development of inference methods for CPs in general (and for survival analysis specifically) is not per se a recent phenomenon. For example, Shiryaev (1963), Hinkley (1970), and Green (1995) introduced and discussed general CP models and Liang et al. (1990) considered the problem of testing the null hypothesis of the absence of CPs. Luo et al. (1997) focused on for CPs in survival data for pre-specified values of t and derived the asymptotic distribution of the partial likelihood ratio test statistic. Recently, Chen et al. (2014) introduced a CP model for detecting a single CP based on Kaplan–Meier estimation of the survival function followed by the least-squares estimation of the CP. Wang et al. (2019) proposed the single CP model for interval-censored survival data with a cure fraction.

The aim of this paper is to extend the REM with CPs (REM-CP) which allows instantaneous changes of the model parameters at certain points in time. Additionally, a statistical test based on Bayes factors (BFs) Jeffreys (1961) is proposed which can be used for quantifying the relative evidence in the data for (non)existence of a CP at a specific time point. The advantage of this approach is that it is also possible to quantify the evidence in the data for nonexistence of a CP. This property is not shared when using frequentist significance testing as a nonsignificant result can either mean 'absence of evidence' or 'evidence of absence' Dienes (2014). Another motivation for using the Bayes factor (BF) is that it is consistent: the evidence accumulates towards infinity for the true hypotheses (either H_0 or H_1 in this case) as the number of observations grows. On the other hand, the classical p value is not consistent as there is always a pre-specified error probability to incorrectly reject H_0 when it is actually true (Type-I error), typically 0.05, even for extremely large samples. To avoid subjective prior specification when computing the BF, we build on a default Bayesian procedure Gu et al. (2018). In addition to testing whether a CP occurred at a given time point, an exploratory CP detection algorithm is proposed which allows researchers to identify CPs in the observational period and thereby create REM-CPs in an automatic and probabilistic principled manner. The CP detection algorithm also relies on BF where CPs are added to the model in the case of very strong evidence in the data, Kass and Raftery (1995), for the existence of a CP. The method we introduce is rather generic and can be applied to other models as well.

Table 2
Overview of important abbreviations and mathematical symbols.

Abbreviations/ mathematical symbols	Description
REM	Relational Event Model
CP	change point
REM-CP	Relational Event Model with CPs
BF	Bayes Factor
X_k	kth statistic
$\boldsymbol{x}_k$	Vector of realizations of the kth statistic
i	event indicator
p	Number of statistics
t	Time point
s	Sender
r	Receiver
D	Event data
H_0	Null hypothesis of nonexistence of a CP
H_1	Alternative hypothesis of existence of a CP
$\beta_k(t)$	Effects (step function) corresponds to the kth statistic
$\beta_k(t'-)$	Effect on the left hand side of a CP in X_k
$\beta_k(t'+)$	Effect on the right hand side of a CP in X_k
$\boldsymbol{\beta}$	Vector of effects
$x_k(t'-)$	statistic corresponding to $\beta_k(t'-)$
$x_k(t'+)$	statistic corresponding to $\beta_k(t'+)$
λ	rate parameters
λ'_t	the summation of the rates over all dyads $\sum_{s,r}\lambda(s,r,t)$
$\mathcal{R}_t$	Risk set
$\mathcal{A}_t$	sequence of events up until time t
γ	Nuisance parameter
$\zeta(t')$	$\beta_k(t-)-\beta_k(t+)$
$\hat{\Sigma}_\zeta$	Fisher information matrix

The paper is organized as follows. In the next section, we describe the REM-CP model. We will discuss two types of applications for our new method: testing for the existence of a CP at a specific (known) time point and the detection of CP(s) on a time scale where the locations of CPs are unknown. In Section 3, we will discuss the implementation of this new model using synthetic data for illustration and testing. In Section 4 we illustrate the use of the REM-CP on real data based on the Apollo 13 voice loops and will test whether and where CPs occurred in during the mission. From a theoretical point of view, our approach fits with the literature on team resilience. Teams require some level of resilience to remain successful, even when faced with circumstances that were not anticipated. In particular, we adopt an approach to team resilience called "graceful extensibility", referring to the way in which (human) systems stretch to handle surprises. In particular, we study how the members of the Apollo 13 relied on a routine where they could, while successfully adapt their interaction dynamics when met with unexpected sudden adversity. The mission members further demonstrate the resilience of the mission by reverting to their original routines where and when possible. This section will provide a more detailed exposition of the aforementioned theoretical perspective. The final section of the paper, Section 5, will summarize the conclusions and discussions. For convenience, Table 2 lists the most important abbreviations and mathematical symbols that we will use throughout this paper.

2. Methodology

In this section, we introduce an extension of the relational event model with change points (REM-CP). A Bayes factor (BF) is described which will be the core building block for testing whether a CP is present or absent at a given time point, which is useful in a confirmatory setup, as well as a CP detection algorithm for finding CPs in a given event sequence and for creating REM-CPs in a principled probabilistic manner, which is useful for exploratory analyses.

2.1. Relational event modeling with change points (REM-CP)

For modeling relational event data as a sequence of tuples (s, r, t), we start with the relational event model (REM) introduced by Butts (2008). REM models the time until the next event by an exponential distribution with rate parameter $\lambda' = \sum_{s,r} \lambda(s, r, t)$, where $\lambda(s, r, t)$, the event rate, is defined over all possible potential dyads (s, r) in the risk set at time t, $\mathcal{R}_t$. The risk set is the set of all combinations of sender and receiver that can transpire at time t. In other words, it is the set of all dyadic events that could be observed at time t. The rate parameter, $\lambda(s, r, t)$, is a piecewise constant function and only changes at the observed event times. We assume that a relational event sequence can be modeled using p predictor statistics $X_1, X_2, \ldots, X_p$ that are functions of the event history until time t, $\mathcal{A}_t$, as well as exogenous information, such as actors' or dyadic attributes (e.g., actors' age or the hierarchical difference between actors). The event history until time τ is given by $\mathcal{A}_\tau = \{(s, r, t) : (s, r) \in \mathcal{R}_t, 0 < t_1 < t_2 < \cdots < \tau\}$. In REM, the event rate between sender s and r at time t is defined as a log-linear function of the statistics given by

$$\log \lambda^{REM}(s, r, t) = \mathbf{x}_{s,r}(t)' \boldsymbol{\beta}. \tag{1}$$

The waiting time until the next event then follows an exponential distribution with rate parameter $\lambda'_t = \sum_{s,r} \lambda(s, r, t)$, which is the sum of the rates all possible potential dyads (s, r) in the risk set at time t, $\mathcal{R}_t$. Furthermore, the probability that the next event pertaining to sender s' and receiver r' at time t follows a multinomial distribution with probability mass function (PMF)

$$p((s, r) = (s', r') | \mathcal{A}_t) = \frac{\lambda^{REM}(s', r', t)}{\sum_{s,r} \lambda^{REM}(s, r, t)} \tag{2}$$

where the denominator is the summation of the rate parameters of all possible dyads that are at risk at time t.

Under the REM-CP, we assume that the vector of coefficients, $\boldsymbol{\beta}(t) = (\beta_1(t), \beta_2(t), \ldots, \beta_p(t))$, $\boldsymbol{\beta}(t) \in \mathbb{R}^p$, is a time dependent p-dimensional vector. Moreover, the coefficients $\beta_k(t)$, $k = 1, \ldots, p$ are step functions associated with statistic X_k at time t. Fig. 1 depicts an example of a step function. The effect for this particular parameter changes at two time points (these are the CPs), namely at $t = 20$ from .1 to .6 and at $t = 40$ from .6 to .4.

The rate parameter for sender s and receiver r at time t under the REM-CP is then defined by

$$\log \lambda^{REM\text{-}CP}(s, r, t) = \mathbf{x}_{s,r}(t)' \boldsymbol{\beta}(t), \tag{3}$$

where the static vector $\boldsymbol{\beta}$ is replaced by the time-varying vector $\boldsymbol{\beta}(t)$. The distributions of the waiting time for the next event and the observed dyad then change accordingly using the rate parameters $\lambda^{REM\text{-}CP}(s, r, t)$. Following the standard REM, the REM-CP also abides the piecewise constant hazard assumption where the rate parameters can only change at the observed event times. This implies that the step functions $\beta_k(t)$ are only allowed to change at the observed event times and not between event times. The testing for (non)existence and of a CP at a given point in time is based on a BF approach, which we introduce in the next section.

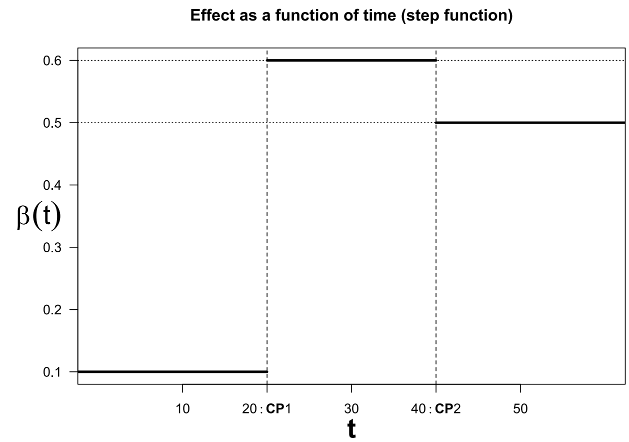


Fig. 1. An example of a step function over time, $\beta(t)$. The y-axis shows the effect size and the x-axis shows the time. The vertical dashed lines show the locations of two CPs.

2.2. Bayes factor CP testing

Testing for the existence of a CP at a specific point in time is, in principle, straightforward. What is required is to quantify the extent to which the data provide more support in favor of a CP than against it. We do this by using the Bayes factor (BF) approach that was introduced by Jeffreys (1961). Using the BF, we quantify the evidence (relative support) in the relational event data for the model with a CP at a given point in time t, as opposed to a model with no CP at that point in time. The proposed CP detection algorithm computes the evidence for a CP for a specific statistic at particular points in time over all potential time points and extends the REM-CP with an additional CP at the time point and for the statistic that results in enough evidence that a CP is present. Before going into detail, we first introduce the CP hypothesis testing using the BF.

We define the two competing hypotheses H_0 and H_1 as follows:

$$H_0 : \text{"time point } t' \text{ is not a CP for statistic } k\text{"} \quad vs.$$
$$H_1 : \text{"time point } t' \text{ is a CP for statistic } k\text{"}. \tag{4}$$

If a CP occurs at time t', the coefficient for β_k should be different *before* t' from its value *after* t'. We define $\zeta_k(t') = \beta_k(t'-) - \beta_k(t'+)$, where $\beta_k(t'-)$ is the β_k parameter before t' and $\beta_k(t'+)$ is the β_k parameter after t'. Hence, $\zeta_k(t') = 0$ implies that there is no CP at time point t', whereas $\zeta_k(t') \neq 0$ implies that there is. The hypothesis given in (4) can be therefore expressed as:

$$H_0 : \beta_k(t'-) = \beta_k(t'+) \quad vs \quad H_1 : \beta_k(t'-) \neq \beta_k(t'+), \tag{5}$$

or alternatively

$$H_0 : \zeta_k(t') = 0 \quad vs. \quad H_1 : \zeta_k(t') \neq 0. \tag{6}$$

In the same way, it is also possible to test whether multiple $\beta_i(t)$'s change instantly at a certain time point t'.

The BF is defined as the ratio of the marginal likelihoods Kass and Raftery (1995):

$$BF_{1,0} = \frac{p(D|H_1)}{p(D|H_0)} \tag{7}$$

where D is the observed relational event data. Hence the BF quantifies how likely the observed event sequence is under hypothesis H_1 (with an additional CP) relative to hypothesis H_0 (without the additional CP). Using equal prior probabilities for the hypotheses (which is a common default choice in Bayesian hypothesis testing), the posterior probability of H_0 and H_1 can be computed as:

$$p(H_0|D) = \frac{1}{BF_{1,0} + 1} \text{ and } p(H_1|D) = \frac{BF_{1,0}}{BF_{1,0} + 1}, \tag{8}$$

Table 3
Interpretation of the BF when testing CPs based on Kass and Raftery (1995).

$log(BF_{1,0})$	$BF_{1,0}$	Interpretation
$0 \leq log(BF_{1,0}) < 1$	1 – 3	Evidence for nonexistence of a CP
$1 \leq log(BF_{1,0}) < 3$	3 – 20	Positive evidence for the existence of a CP
$3 \leq log(BF_{1,0}) < 5$	20 – 150	Strong evidence for the existence of a CP
$log(BF_{1,0}) \geq 5$	≥150	Very strong evidence for the existence of a CP

which quantifies the probability that there is no CP at time t' given the observed data, and the probability that there is a CP at time t' given the observed data.

To avoid the need to manually specify the prior based on external prior knowledge, a fractional Bayesian approach is considered where the prior is automatically constructed using a minimal fraction from the data O'Hagan (1995), so that maximal information is used for quantifying the evidence in the data between the hypotheses. Furthermore, by centering the prior at the null value, positive and negative effects are equally likely a priori Mulder (2014). In a final step, we adopt the proposal of Gu et al. (2018) using Gaussian approximations of the posterior,[1] resulting in a so-called Savage–Dickey density ratio (Dickey (1971), Wetzels et al. (2010) and Gu et al. (2018)) which can be computed straightforwardly from a standard (classical) analysis of the model,

$$BF_{1,0} \approx \frac{\pi_1(\zeta_i = 0)}{\pi_1(\zeta_i = 0|\boldsymbol{D})} = \frac{N(0|0, n\hat{\Sigma}_\zeta)}{N(0|\hat{\zeta}, \hat{\Sigma}_\zeta)}, \tag{9}$$

where $N(0|m, \Sigma)$ denotes the normal density with mean m and covariance matrix Σ evaluated at 0, and where the posterior is approximated with a normal distribution centered at the MLE, $\hat{\zeta}$, with covariance matrix $\hat{\Sigma}_\zeta$, which corresponds to the error covariance matrix of the MLE, i.e., the inverse of the Fisher information matrix, and the fractional prior contains the information of a single observation (by rescaling the error covariance matrix with the sample size n) with mean 0 (so that positive effects are equally likely a priori as negative effects).

To facilitate the interpretation, Kass and Raftery (1995) proposed a rule of thumb to group BFs into different categories that classify the strength of the evidence in favor of a hypothesis (see Table 3). For example, there is "positive" evidence for H_1 when the BF lies between 3 and 20 and, equivalently, positive evidence for H_0 when the BF lies between 1/20 and 1/3. These labels provide some rough guidelines when interpreting BFs which may be useful for researchers who are relatively new to the methodology. Even though we follow these interpretations in this paper, note that interpretations may differ depending on the problem under consideration such as extant of genuine prior knowledge, experimental conditions, and external stimuli. The guidelines can be useful as a classifier to provide a rough boundary for interpreting the relative evidence as provided by the BF. See also Jeffreys (1961) and van Doorn et al. (2021) for more elaboration on the interpretation of BFs.

In this paper, we use this BF approach (i) to test whether a CP is present at a given time point and (ii) to detect CP(s) on the entire time scale. In the next two sections, we will show how we cover both uses by means of Eq. (9) for the observed relational event data.

2.2.1. Testing for the existence of a CP at a given time point

A researcher may want to test whether a CP occurs at a specific point in time and how much evidence there is to support the (non)existence of a CP at that given time point. For example, when studying communication patterns in a cardiovascular surgery team, a researcher might be interested in testing whether the communication of the surgery team changed at times when the patient shows an unexpected response (e.g., a decreased heart rate and sudden low blood pressure). The proposed Bayes factor CP detection algorithm has been developed to test whether such specific points in time indeed show CPs in the communication dynamics of the team. In Table 4, we show the pseudo-code for the BF algorithm for CP detection at a certain point in time.

2.2.2. CP detection algorithm

If a researcher does not have expectations when exactly a CP occurs but it is expected that one or more CPs occurred, the CP detection algorithm is useful. In this case, the question of interest becomes whether there are abrupt changes in interaction dynamics in the data and, if so, when they occur and how the interaction dynamics changed at those time points. In this section, we will introduce a step-by-step CP detection algorithm that can find CPs across the complete event history.

In the first step, we specify the statistics we want to include in our REM (see also Butts (2008)) and we specify a grid of time points in the observational period where CPs potentially could be observed. This grid could either be fine or rough (we come back to this later). Second, we apply the steps in Table 4 to compute the evidence of a CP at each time point and for each statistic. Third, we identify the statistics and time point that has the largest evidence over all statistics and time points and/or for which the evidence for the existence of a CP exceeds a threshold. We call this a "detected CP". Then, we extend the REM-CP with an additional CP. This implies replacing the original statistic with two new ones. The first contains the original values of the statistics until the CP, and zeros elsewhere and the second statistics contain zeros until this detected CP and the original values of the statistic afterwards. We repeat this cycle of testing for CPs and splitting detected CPs until no more CPs achieve sufficient evidence. We provide a detailed overview of the steps in Table 5 and a general flowchart in Fig. 2.

Throughout this paper, the CP detection algorithm is applied using a threshold value of $B_{cutoff} = 150$, which implies "very strong" evidence for the existence of a CP (Table 3). This is a fairly conservative threshold which will ensure that the model is only extended in the case of very clear evidence that a CP is present. In the following section, we show that this operationalization of the algorithm results in a desirable behavior. We come back to this choice in the Conclusion section.

3. Synthetic data analysis

In this section, we illustrate the performance of the step-by-step CP detection algorithm using synthetic data. For this purpose, we generated data considering the following three different scenarios:

- **Scenario 1**: One statistic has a CP and the others have no CP.
- **Scenario 2**: All statistics have a CP at the same time point.
- **Scenario 3**: Each statistic has a CP and all CPs occur at different points in time.

Each data set consists of 15 actors and 5000 relational events. We include two endogenous and two endogenous statistics into the model. As we use the synthetic data for illustrative purposes (rather than as a simulation of all possible scenarios), we choose three straightforward scenarios. The endogenous statistics in scenario 1 included transitivity and inertia, in scenario 2 these were inertia and outdegree receiver and in scenario 3 these were reciprocity and outdegree receiver. The two exogenous statistics can be, e.g., age, sex, location, and hierarchy of senders/receiver and are constant over the network. Here, we simulated location and hierarchy. We considered three locations and twelve hierarchical levels. For each scenario, we generated 30 relational event datasets for a total of 90 synthetic datasets. In all scenarios we generated data by repeating the steps in Table 6:

When following these steps, the resulting synthetic data have a form that is similar to the empirical relational event sequence in Table 1 (except that the simulated events do not have any content, but note that the content of the messages is also not used in the relational event analysis of the Apollo 13 data in this paper).

[1] Note that the Bayesian information criterion (BIC), Raftery (1995) and Schwarz (1978), which is extensively used in statistical practice, also relies on normal approximations of the kernel of the posterior.

Table 4
Step-by-step methodology for testing the (non)existence of a CP testing at a certain point in time t'.

Testing for the (non)existence of a CP at a given time point	
1.	Specify a time point of interest, say t', as well as a statistic, say X_k, under a REM for which we want to test whether a CP occurred.
2.	Replace the effect $\beta_k(t')$ with two effects, $\beta_k(t'-)$ and $\beta_k(t'+)$, and with corresponding statistics $x_{k_{t'_-}}$ and $x_{k_{t'_+}}$, where $x_{k_{t'_-}}$ contains the original values of x_k until time point t', and zeros elsewhere, and $x_{k_{t'_+}}$ contains zeros until time point t' and the original values of x_k after that. The other statistics remain unaltered.
3.	Fit a REM using the new data set and the expanded set of parameters and extract the estimates and error covariance matrix for the parameters $\beta_k(t-)$ and $\beta_k(t+)$. Denote the mean and covariance matrix by $\boldsymbol{\mu}_{t'}$ and $\boldsymbol{\Sigma}_{t'}$.
4.	Approximate the posterior of the parameter of interest, $\zeta(t') = \beta_k(t-) - \beta_k(t+)\|D \sim N(\mu_\zeta(t'), \sigma^2_{\zeta(t')})$, where $\mu_\zeta(t') = \mu(t'-) - \mu(t'+)$, $\sigma^2_{\zeta(t')} = [1\ \ -1]\boldsymbol{\Sigma}_{t'}[1\ \ -1]^T$ and $\mu(t'-)$, $\mu(t'+)$ are means of $\beta_k(t-)$ and $\beta_k(t+)$ respectively, extracted from $\boldsymbol{\mu}_{t'}$.
5.	Compute the evidence for the existence of a CP at time t' as the ratio of the default prior of $\zeta(t')$ at zero and the posterior of $\zeta(t')$ at zero (see Eq. (9)).
6.	Interpret this evidence according to Table 3 and decide to extend the model with an additional CP.

Table 5
Pseudo-code for the CP detection algorithm.

CP detection algorithm	
1.	Specify a REM with a set of endogenous and exogenous statistics of interest.
2.	Create a grid of possible CPs over the observed event history, i.e., $t'_1, \cdots, t'_G$.
3.	For $k = 1, \cdots, P$ (across all statistics) and for $g = 1, \cdots, G$ (across all grid points), compute the relative evidence in the data between the following hypotheses $H_1 : \beta_k(t'_g-) \neq \beta_k(t'_g+)\ vs\ H_0 : \beta_k(t'_g-) = \beta_k(t'_g+)$, using the steps from Table 4.
4.	Check for which time point t'_g and for which statistic k the evidence for H_1 (i.e., evidence in favor of a CP at t'_g for statistic k) is largest.
5.	If the largest evidence for the existence of a CP from the previous step exceeds a threshold B_{cutoff}, the model is extended with a CP at that time point and for that statistic by splitting the vector $\boldsymbol{x}_k$ into $\boldsymbol{x}_k(t'_g-)$ and $\boldsymbol{x}_k(t'_g+)$ as described in Section 2.2.1. The number of statistics P is increased with 1.
6.	Repeat the algorithm (from step 3) and continue until the evidence for the existence of a CP is no longer larger than the threshold value for all time points and all statistics.

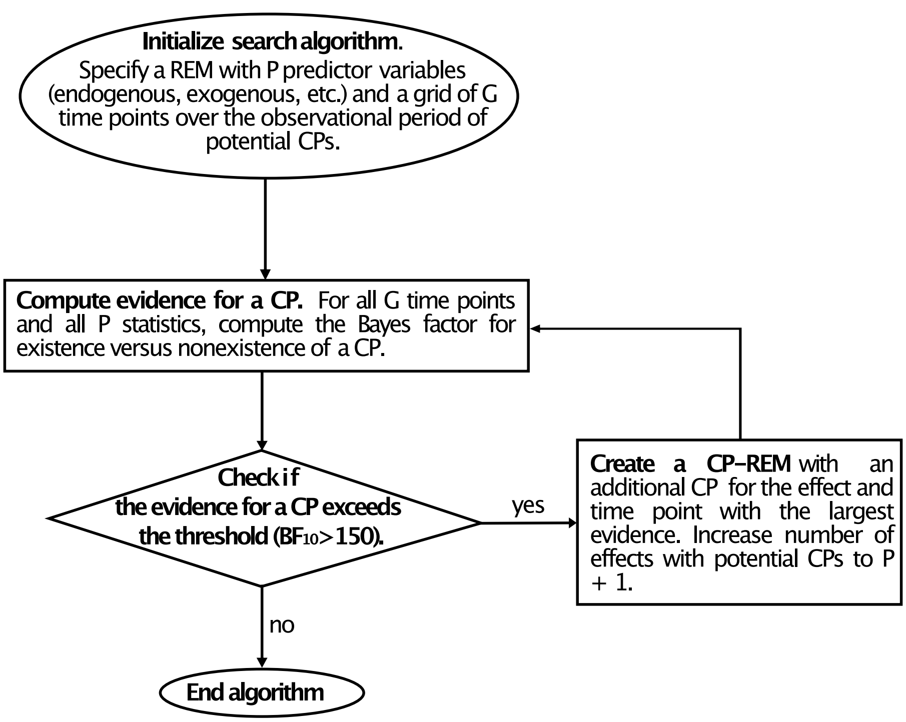


Fig. 2. A flowchart with the key steps of the CP detection algorithm (see also Table 5).

3.1. Scenario 1: One statistic has one CP and the others have no CP

In this scenario, we impose one CP on inertia at the location of $t = 5$ and no CP for the other statistics. The CP detection algorithm is applied from Table 5 using a threshold for the BF of 150 (i.e., "very strong" evidence for the existence of a CP). The y-axis of Fig. 3 contains the log BF for the existence of a CP (computed from Eq. (9)) and the x-axis shows the time. The first row shows that the highest evidence for a CP occurs for inertia at $t = 5$–there are no other statistics and no other time points that have a higher evidence for a CP. Even though the data were generated with only a CP at time point $t = 5$ for inertia, there is also some evidence for a CP for transitivity and the two exogenous statistics.

Table 6
Steps for data generation.

Generating relational events under a REM-CP	
0.	Initialize the REM-CP with zeros for the endogenous statistics and compute the rate parameters for all dyads at time 0, i.e., $\lambda(s, r, 0)$.
1.	Sample the event time $t_m = t_{m-1} + \Delta t_m$, where $\Delta t_m \sim \text{Exponential}(\sum_{s,r} \lambda(s, r, t_{m-1}))$.
2.	Sample the next dyad from the categorical distribution in Eq. (2).
3.	Update the endogenous statistics for the next event.
4.	Update $\beta(t)$ after a CP if present.
5.	Update the event rate for all dyads for the next event using formula (3).
6.	Repeat steps 1 to 5 until 5000 events have been generated.

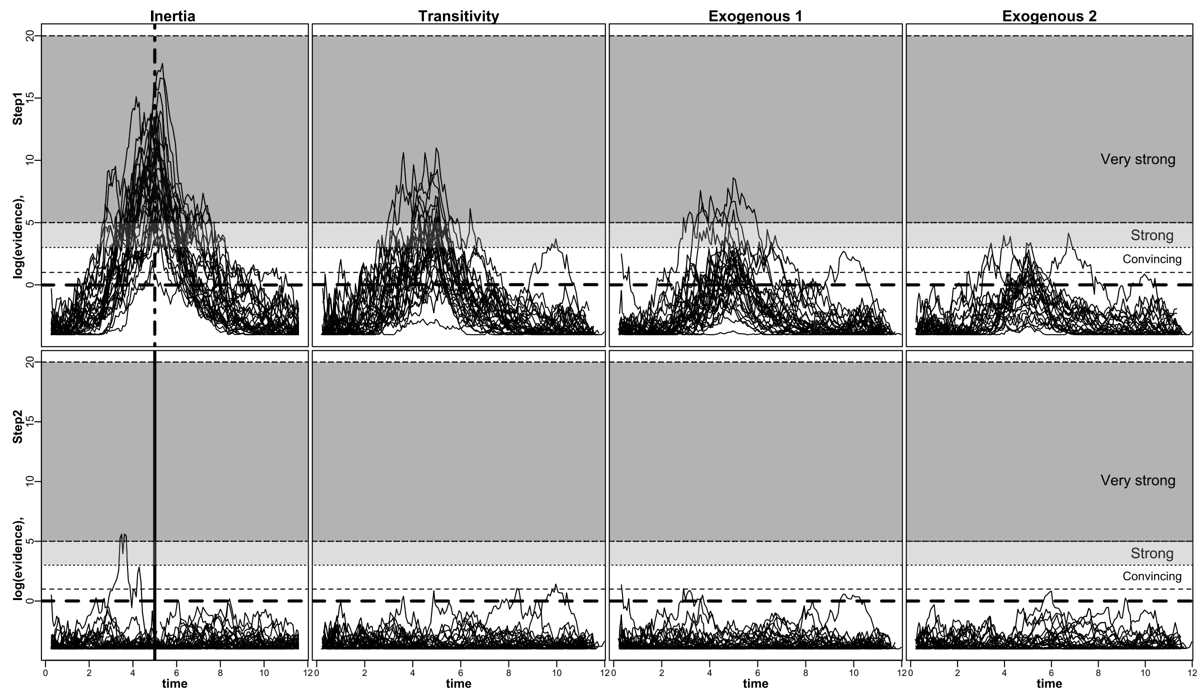


Fig. 3. Scenario 1: Inertia has one CP and the other statistics have no CP. The y-axes contain the log evidence of the existence of a CP computed from Eq. (9), and the x-axes include the event time. The dark gray area corresponds to "very strong" evidence, while the light gray and white areas show the "strong" and "positive" evidence, Kass and Raftery (1995), respectively, as indicated to the right of each row of plots. The vertical solid black line indicates the detected CP in each step. The algorithm continues until the results denote there is no evidence for the existence of any further CPs.

This is caused by the fact that there was a change in interaction dynamics at $t = 5$ and inertia affects the communication dynamics overall. However, the evidence is largest for inertia. Following the algorithm, we specify a REM-CP with a CP for inertia at $t = 5$ and, controlling for this CP being there, examine further evidence for other CPs. We fix the detected CP of inertia in step 1 on its inferred location and create two statistics (as explained in Section 2.2.1). Hence, instead of the four statistics from our original REM, Butts (2008), there are now five statistics in the REM for the second step.

The second row of Fig. 3 shows the results of the second step. For clarity of exposition, we merged the plots for the two inertia statistics into a single plot and show them in the bottom-left panel of Fig. 3. The vertical solid black line indicates the location of the detected CP. As all panels show, there is no convincing evidence for any additional CP, now that we control for inertia having a CP at $t = 5$. Therefore, the algorithm is stopped resulting in a single REM-CP with a single CP at $t = 5$ for inertia. Also note that there is no convincing evidence anymore for a CP for transitivity, exogenous 1, and exogenous 2 now that we have fixed the CP for inertia after step 1. Indeed, the evidence for a CP for those statistics in step 1 appears to have been triggered by inertia changing value at that point in time.

For visual clarity, in all the following scenarios we will merge the plots for the new statistics inside the panel of the corresponding statistic, like we did for inertia here.

3.2. Scenario 2: All statistics have the same known CP

Scenario 2 imposes a single CP on the same location, $t = 14$, for all statistics. That is, the locations of CPs are the same for all statistics. Fig. 4 shows the results of our algorithm for Scenario 2. In the top row, it can be seen that "Exogenous 1" has the highest evidence for a CP and this is found on the right location, $t = 14$. This time point is the first detected CP.

The second row shows the evidence for the CPs after fixing the CP for "Exogenous 1" by splitting the statistic into two. Visually, they are again merged into the same panel and the detected CP is highlighted by a solid vertical line for that statistic at $t = 14$. The maximum evidence in the second row is found for "Exogenous 2", again far above the threshold, Kass and Raftery (1995), also at $t = 14$. In the third step of the algorithm, both CPs are fixed and the maximum evidence occurs at $t = 14$ for inertia. Following the algorithm, the fourth step/row finds sufficient evidence for a CP for Outdegree Receiver at $t = 14$.

Because statistics can potentially change at multiple points in time, we continue the search for CPs in row five. In this row we can see that there is no convincing evidence, Kass and Raftery (1995) for any other CPs in the data. Therefore, there is no need for further steps and we terminate the algorithm at this step. The conclusion is that all statistics have a CP at $t = 14$ and that the model detected all CPs correctly.

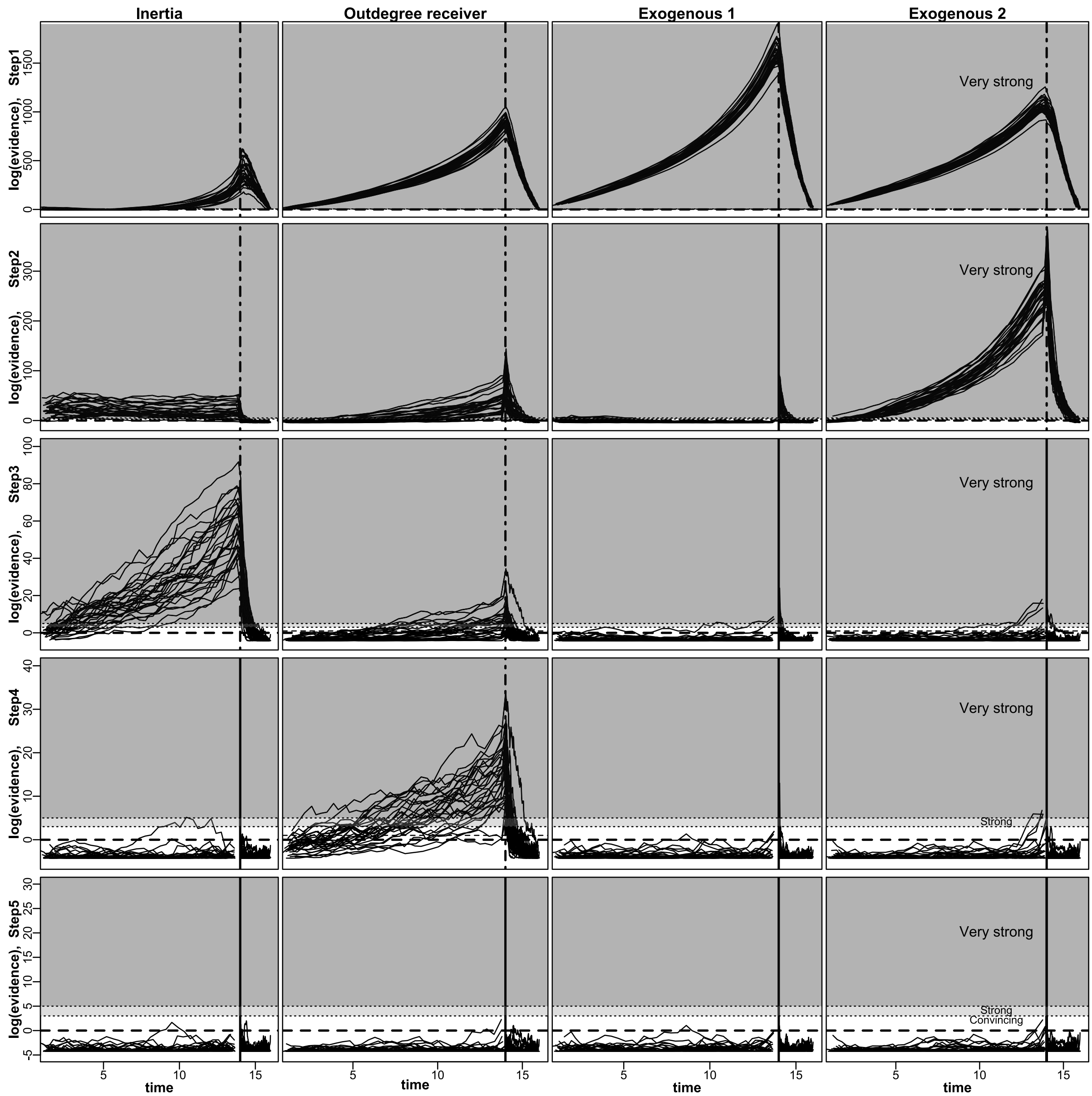


Fig. 4. Scenario 2: All statistics have the same CP locations. The y-axis corresponds to the log Bayes factor for the existence of a CP computed from Eq. (9) and the x-axis shows the event time. The dark gray area corresponds to "very strong" evidence while the light gray and white areas show the "strong" and "positive" evidence, Kass and Raftery (1995), respectively, as indicated to the right of each row of plots. The vertical solid black line indicates the detected CP in each step. In this scenario, the proposed model detected all the CP in about their actual places, $t = 14$. The algorithm continues until there is no evidence for the existence of any additional CP.

3.3. Scenario 3: Each statistics has its own CP and all CPs are dissimilar

The third scenario is a more challenging scenario where the locations of the CPs are different for each statistic. We generated data in which the CPs for reciprocity, outdegree receiver, exogenous 1, and exogenous 2 are located at $t = 8$, $t = 6$, $t = 4.5$, and $t = 3$, respectively.

The various steps are shown in Fig. 5. The first row shows that the maximum evidence for a CP is for "Exogenous 2" at $t = 3$.

In the second step (second row) the largest evidence is found for a CP at $t = 6$ for outdegree receiver (although reciprocity has only slightly less high evidence for a CP in this step). The third and fourth rows reveal CPs for "Exogenous 1" at $t = 4.5$ and reciprocity at $t = 8$. No sufficient evidence is found for additional CPs. The CPs in the data are retrieved by the algorithm, highlighting the way in which the approach works.

4. Empirical data analysis

4.1. Apollo 13 data

In our empirical analysis, we study the dynamics of the Apollo 13 mission around the time of "Houston, we've had a problem". From a theoretical point of view, our approach fits with the literature on team resilience. Workplace team resilience has been proposed as a potential asset for work teams to maintain performance in the face of adverse events Hartwig et al. (2020). Team resilience has been defined in many ways, but the most fitting with the analysis at hand is that of "graceful extensibility", a concept that refers to the way in which (human) systems stretch to handle surprises (Woods and Branlat (2011), Woods (2015)). Systems with finite resources in changing environments are always experiencing and stretching to accommodate events that challenge boundaries. Without some capability to continue to stretch in the face of events that challenge boundaries, systems are more brittle

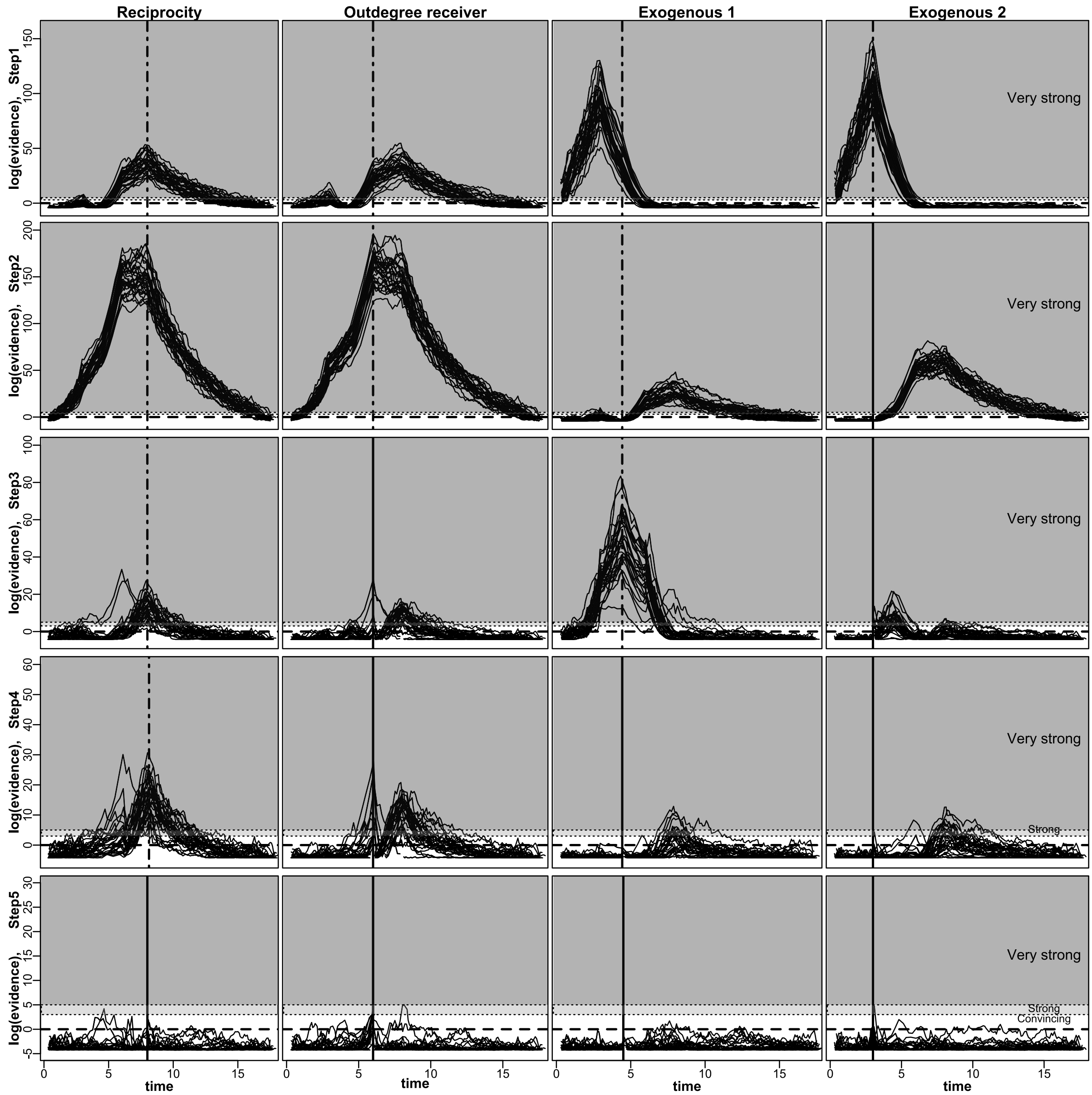


Fig. 5. Scenario 3: Each statistic has its own CP and all CPs are dissimilar. The y-axis shows the log Bayes factor for the existence of a CP computed from Eq. (9) and the x-axis indicates the event time. The columns refer to statistics and each row refers to one step from step by step algorithm which is indicated in the y-axis. The dark gray area corresponds to "very strong" evidence while the light gray and white areas show the "strong" and "positive" areas, Kass and Raftery (1995), respectively, as indicated to the right of each row of plots. The vertical solid black line indicates the detected CPs in each step. In this scenario, the proposed model detected all the CPs in their actual places.

than stakeholders realize and all systems, however successful, have boundaries and experience events that fall outside these boundaries (ie. the surprises) Woods (2015). There is always some rate and kind of events that occur to challenge the boundaries of more or less optimal or robust performance, and thus graceful extensibility, being prepared to adapt to handle surprise, is a necessary form of adaptive capacity for all systems Woods (2006).

The general idea behind team resilience is that teams require some level of agility of flexibility to remain successful, even when faced with circumstances that were not anticipated. This is an idea that has been studied across many types of teams, from business teams to sports teams to emergency response teams to space crews. Where teams adopt routines to operate as optimally as possible under non-surprising conditions, the main challenge is whether they are able to reconfigure and improvise under unforeseen circumstances Janssens et al. (2022). Two aspects are important when considering resilience and graceful extensibility in the context of human teams.

First, teamwork, especially the kind of teamwork where the team needs to solve problems building on the varied expertises and skills of its members, is essentially built on the communication between its members Kratzer et al. (2006). Only through their communication can they jointly come to an organized coordinated response and adapt their activities appropriately in the face of unexpected emergencies. The result of this is that it is essential to study such teams as communication networks, where the ability of the team to adapt its network appropriately creates the foundation for its potential for graceful extensibility (cf., Janssens et al. (2022)). Our study is not the first to take a communication network approach to space missions; in fact, Larson et al. (2019) rephrase the famous Apollo 13 citation to "Houston, we have a teamwork problem".

Second, in their scoping review of the team resilience literature, Chapman et al. (2020) conclude that there has been a reliance on cross-sectional research designs in empirical studies of team resilience, which is fundamentally incongruent with the dynamic nature of the concept. Teams can only be resilient to the extent that they are able to *adapt* and *change* their joint behavior, which means that the study of resilience requires a researcher to be able to compare the "normal" communication and coordination flow to those that occur once an adverse event has happened. Ideally, one also studies how the team moved to this new state or how it failed getting there.

Taken together, understanding the (lack of) the ability of teams to continue to function well in the face of adverse surprises requires a focus on the temporal dynamics of the communication/coordination interaction network among the team members. In this study, we adopt this focus, using a relational event CP model. The model allows us to see how and when the Apollo 13 mission members adapted their communication when the main problem occurred and how they altered their communication and coordination in order to take care of the main problem and attempt to move back to the patterns at which they are the most comfortable and that get as close as possible to their "routine" interaction at which they have been trained the most.

We note that ours is not the first study in which the REM is used to unravel how teams respond to emergencies. Butts (2008) studied the communication networks of emergency responders after the World Trade Center disaster occurred on September 11, 2001, when two hijacked airliners were flown into the North and South Towers of the WTC complex. The main difference between the two approaches is that our study uses CPs, allowing the communication to change over time, whereas Butts (2008) fitted a single model across the observation period.

4.2. Apollo 13 data preprocessing

In our empirical example, we analyze part of the Apollo 13 mission. The mission launched as scheduled at 2:13:00 EST (19:13:00 UTC) on April 11, 1970. On board were James Lovell Jr. (Commander, CDR), John "Jack" Swigert Jr. (Command Module Pilot, CMP), and Fred Haise Jr. (Lunar Module Pilot, LMP). The mission was already the seventh crewed Apollo mission and the third to land on the moon.

The main parts of the Apollo 13 were the Command and Service Module (CSM) Odyssey and the Lunar Module (LM) Aquarius. At 3 h and 6 min (003:06) after launch, the CSM was separated from the Saturn-V launch vehicle and docking with the LM occurred at 003:19, all within seconds from the scheduled moments. In fact, the mission went nicely according to plan and the crew entertained live television broadcasts during considerable parts of the trip. In fact, also the mid-course correction burn (after 30 h:13 m into the mission) was broadcast live. Overall, the mission was quite routine and everything went to plan, except when Swigert suddenly realized he had forgotten to file his federal income tax return that would be due during their trip and mission control had to check if he could get an extension for it.

At almost 56 h into the mission, Sy Liebergot (Electrical, environmental, and consumables engineer, EECOM) at mission control requested that Swigert would stir the contents of oxygen tank 2. Ninety-five seconds after the stir, at 56:54:53, the astronauts heard a "pretty large ban" and experienced fluctuations in electrical power and control thrusters. This set a series of events in motion. While the oxygen sensors show that the oxygen supply is decreasing, Lovell notices, when looking out the window, that Odyssey is leaking oxygen into space. With oxygen levels depleting fast, the astronauts not only faced a risk of running out of oxygen to breathe but also suffered from the insufficient voltage (since the fuel cells required oxygen to generate the voltage) to maintain almost all equipment in the CSM.

Our change point analysis focuses on the events from 54:46:28 to 62:06:53 (hh:mm:ss). This sequence starts about an hour before the incident and ends with the message given: "Apollo 13 is now safely back on a trajectory towards Earth, with a stable configuration and no immediate dangers". The data come from the Apollo 13's voice loops transcripts, obtained from Apollo 13 Real-time, Tseng (2017) and Apollo Flight Journal (2015); the data include the Flight directors' voice loop and the air-ground's voice loop. Flight directors (Houston's Mission Control Center) were located in Houston and the crew (astronauts) were connected to this control center via Capsule Communicator (CAPCOM).

The Apollo 13 data provide an ideal illustration for our method since the mission clearly moved quite suddenly from a routine endeavor into to a critical mission- and life-saving operation where both the astronauts and mission control had to solve unexpected and urgent problems in order to bring the crew home alive. Indeed, van den Oever and Schraagen (2021) found that the communication pattern in Apollo 13 Flight director's voice loop after the incident had changed. Communication during a space mission is highly protocolized. In particular, the only person at mission control who was able/allowed to address the crew is the Capsule Communicator (CAPCOM). The reason for this is that if everyone at mission control would be allowed to talk to the crew members, this could lead to noisy, unclear, and confusing communication.

Our analysis starts an hour before "the incident". Considering the routine nature of the mission up to the incident, we expect this hour of events to represent the "normal state" of the communication; this state is likely to change after the incident. There were some commercial gaps in the transcripts before the incident and no data are available of what was mentioned during these breaks. Thus, we have some periods with no observed events. To make the event data homogeneous, we replaced these gaps with a moving average of waiting times before each gap.

The final data include 19 actors (three astronauts and sixteen members at mission control) and 5498 events in about eight hours. The full elapsed time is from 54:46:28 to 62:06:53 (hh:mm:ss), with the oxygen tank incident occurring at 55:54:53. The aggregated communication network of Apollo 13 data over the entire event history is shown in Fig. 6 where nodes reflect the actors and arrows correspond to ties between actors (s, r). Table 7 denotes the list of actors in the data and the first few records of social interactions observed in Apollo 13's voice loops data are shown in Table 1.

4.3. Results of the apollo 13 analysis

First, we apply a conventional REM (ie. without any CPs) to the data. Next, we test for the existence of a CP around the time of the incident. Finally, we analyze the entire event sequence to detect CPs over the entire observational period using the proposed algorithm.

4.3.1. REM without CPs

In this subsection, we apply the conventional REM model on Apollo 13 data. The statistics included in the model are indegree sender, reciprocity, outdegree receiver, same location (ie. are sender and receiver both at mission control/both in space, or is this a cross-location communication), fixed effects for TELMU, CDR, LMP, CAPCOM, EECOM, and FLIGHT, see Table 7. The output is summarized in Table 8.

The conventional REM presumes that interaction behavior between actors is static and, therefore, does not alter over time. Under this restricted assumption, all effects are significant (according the computated p values) and the posterior probability of a nonzero effect is also approximately 1 for all effects. Given all other effects in the model, the effect of indegree sender suggests a higher propensity for actors to initiate contact/communication in the future who have received more past communication. This makes sense, as this fits with the back-and-forth that occurs when solving a complex problem. Although the effect corresponding to reciprocity is small, it is significant and depicts the propensity to reciprocate past interactions. Outdegree of the receiver effect is negative, which implies that actors who sent a lot in the past are less likely to be receivers in the future. The statistic "same location" refers to an exogenous actor attribute that affects dyad's rate of interacting based on whether sender and receiver have the

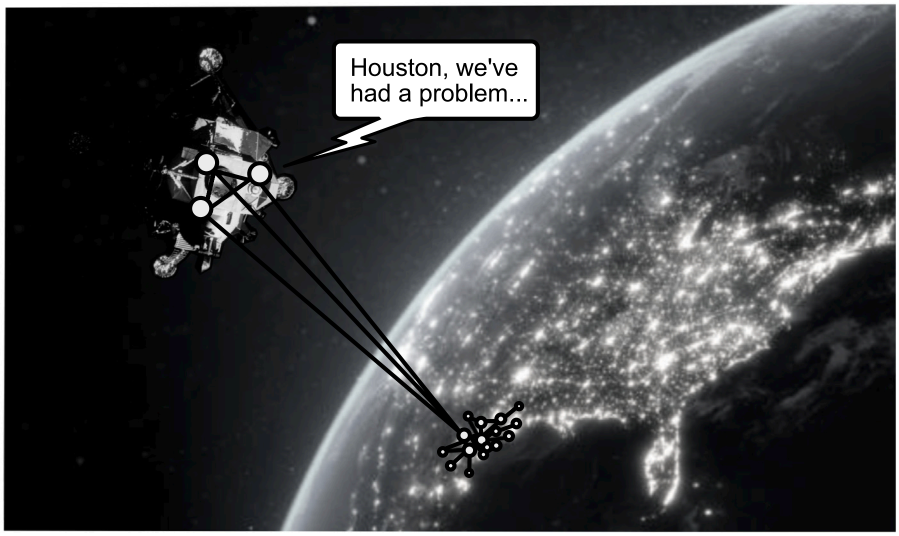


Fig. 6. Aggregated network of Apollo 13's voice loops data over the entire event history.

Table 7
Actors in Apollo 13 data.

Acronym of actors	Meaning	Team
AFD	Assistant Flight Director	Flight directors
CAPCOM	Capsule Communicator	Flight directors
CONTROL	Control Officer	Flight directors
EECOM	Electrical, Environmental and Consumables Manager	Flight directors
FAO	Flight activities officer	Flight directors
FLIGHT	Flight Director	Flight directors
GNC	The Guidance, Navigation, and Controls Systems Engineer	Flight directors
GUIDO	Guidance Officer	Flight directors
INCO	Integrated Communications Officer	Flight directors
NETWORK	Network of ground stations	Flight directors
PROCEDURES	Organization & Procedures Officer	Flight directors
RECOVERY	Recovery Supervisor	Flight directors
RETRO	Retrofire Officer	Flight directors
SURGEON	Flight Surgeon	Flight directors
TELMU	Telemetry, Electrical, and EVA Mobility Unit Officer	Flight directors
CDR	Commander James A. Lovell Jr.	crew (astronauts)
LMP	Lunar module pilot Fred W. Haise Jr.	crew (astronauts)
CMP	Command Module Pilot John (Jack) L. Swigert Jr.	crew (astronauts)

Table 8
Results of relation event modeling without CPs.

Sufficient statistic	MLE	Standard error	p-value	$P(\text{nonzero effect}\|data)$
Indegree Sender	0.72725	0.01132	0.00	1.00
Reciprocity	0.18119	0.00602	0.00	1.00
Outdegree receiver	−0.46805	0.02201	0.00	1.00
Same location	−1.23369	0.05096	0.00	1.00
FETELMU	4.43333	0.06582	0.00	1.00
FECDR	3.71787	0.08234	0.00	1.00
FELMP	3.73044	0.07837	0.00	1.00
FECAPCOM	4.39497	0.05138	0.00	1.00
FEEECOM	4.22553	0.06455	0.00	1.00
FEFLIGHT	6.10804	0.07621	0.00	1.00

same value (or not) on this attribute. That is, whether the sender and receiver belong to the same group: Flight directors (mission control team) and crew (astronauts) which shows a negative effects. For actors whose situation (position, location, etc.) affects their communication pattern, Butts (2008) suggests to incorporate fixed effects into the relational event model. All fixed effects are positive throughout the event sequence. This indicates the specific tendencies for these actors to participate in communication (controlling for the other statistics in the model).

4.3.2. Testing the existence of a CP around the time of "Houston, we've had a problem"

At $t = 0.33$ (transcript time 55:55:20) the commandar (CDR), James Lovell Jr., said "I believe we've had a problem here" and a few seconds later he emphasized "Houston, we've had a problem. We've had a MAIN B BUS UNDERVOLT". This phrase has become famous and marks the onset of an unanticipated issue which put the Apollo 13 mission in a crisis and started a critical situation for both mission control team and astronauts. This suddenly changed the mission from a fairly routine trip to the moon to the mission being aborted and focus on saving the lives of the Apollo crew. Hence, this time point has the potential to be a CP that alters the dynamics of the mission's communication: changing from "business-as-usual" to crisis-mode problem-solving. In this subsection, we hence explore the existence of a CP at $t = 0.33$ (55:55:20) for a set of statistics of interest such as indegree sender, outdegree receiver, reciprocity, same location, and fixed effects of FEFLIGHT, CAPCOM, TELMU, CDR, LMP, EECOM. Table 9 shows the results. Indeed, we find strong evidence that the incident with the oxygen tank caused a sudden change in the interaction dynamics among the members of the mission. In particular, we observe that the CP occurred for indegree sender, Outdegreer receiver, reciprocity, same location, FECAPCOM and FEFLIGH for which the conventional REM could not be sufficient to catch them. However, that should be noted that not every aspect of the interaction dynamics changes at this point, in the data there is

Table 9
Results when testing for the (non)existence of a CP at 55:55:20 (hh:mm:ss). The first column depicts the statistics and the second column shows the evidence (ie, the logged BF) for the CP at this point in time. The third and fourth columns provide the posterior probability of a CP at this time point and its interpretation according to Table 3. The fifth and sixth columns represent the coefficient before and after the corresponding CP.

Time point: 55:55:20					
Statistic	$log(B_{1,0})$	P(CP exists$\|data$)	Interpretation of the evidence	Coefficient before CP	Coefficient after CP
Indegree Sender	253.35	0.999	Very strong for a CP	0.84	0.08
Outdegree receiver	139.21	0.999	Very strong for a CP	−0.99	−0.29
Reciprocity	258.39	0.999	Very strong for a CP	−0.07	0.20
FEFLIGHT	97.24	0.999	Very strong for a CP	4.89	7.46
Same location	64.90	0.999	Very strong for a CP	−2.89	−1.36
FECAPCOM	8.21	0.890	Very strong for a CP	3.81	4.40
FETELMU	−4.03	0.017	Strong for no CP	4.10	4.44
FECDR	77.076	0.999	Very strong for a CP	1.82	4.60
FELMP	65.86	0.999	Very strong for a CP	1.70	4.38
FEEECOM	−3.90	0.020	Strong for no CP	4.62	4.21

evidence that the effect of TELMU remains unaltered around the time of the incident. We will discuss this further in Section 4.3.3.

4.3.3. Detecting CP(s) on the entire Apollo 13's voice loops data

For a researcher, the incident with the oxygen tank highlights an obvious point in time where it is very likely that the crew and flight director staff would have to change their communication flow. In most data sets, researchers may not have the luxury of the data already shouting "change point!" to them. In this case, the logical approach is to explore whether CPs occur at some points in the data, rather than pre-specifying which time point to consider specifically. In fact, even the communication patterns in the Apollo 13 data may have additional points of change that do not have an onset as dramatic as an exploding oxygen tank. Therefore, we now illustrate the approach of building REM-CPs without any assumptions about where CPs might be present. That is, we assume that the locations of CPs are unknown and should be detected from the data. Hence, we apply our "REM-CP" detection algorithm on the whole Apollo 13 voice loops event history data.

The result of this phase of our study is displayed in Fig. 7: for each statistic (in the columns), the figure shows the evidence for a CP at any point over the event sequence. As explained in Section 2.2.2, we follow a stepwise approach, in which we consider the highest evidence for a CP across all statistics and all time points and, if that evidence is large enough to be considered strong or very strong, fix that CP (ie. a CP for one specific statistic at that specific point in time) in our model. Then, we repeat this in the next step, while controlling for the previously detected CP(s). This is repeated until there no longer is a CP that shows strong evidence of being an additional CP. Fig. 7 shows all 21 steps in the rows, from top to bottom. The figure is included as a visual illustration of the method. In larger models, such a plot may become unwieldy. However, the algorithm does not require a plot. The plot is mainly of use for illustration of the algorithm and can aid the interpretation of the researcher.

Fig. 7 shows that the most evidence for a CP for any statistic at any point in time is for *indegree sender* around the time of the incident. In fact, the effect of *indegree sender* changes two CPs with very strong evidence very close to each other at that time. In fact, step 6 shows another CP for this statistic around that same time. Rather than interpreting all CPs separately – although this would be a very interesting thing to do for a micro-analysis of exactly how the communication dynamics change after the incident – we suggest to focus on the bottom row of Fig. 7 which shows all of the detected CPs (and indicates that no additional CP has enough evidence). This bottom row is also plotted in the left column of Fig. 9. From the location of the vertical solid black lines in each panel, we can see that the CPs occur around three points in time: around t = 0.33 (55:55:10), t = 0.42 (58:04:50), and t = 0.52 (60:28:50).

The first CP area is very close to the time point where the Apollo 13 commander announced "I believe we've had a problem here" (55:55:20) and a few seconds later followed up with a stronger "Houston, we've had a problem. We've had a MAIN B BUS UNDERVOLT". This CP is visible in reciprocity, indegree sender, FLIGHT, CDR, EECOM, and same location. The flight director Eugene Kranz (FLIGHT) was leading the ground team and was the most active sender and receiver. Flight directors during Apollo had a one-sentence job description, "The flight director may take any actions necessary for crew safety and mission success" (Wikipedia contributors (2022)). FLIGHT had the most authority for the duration of the flight and monitored the other flight controllers, remaining in constant communication with them via intercom channels. EECOM was in charge of all Electrical, Environmental, and Communications systems. Because oxygen tanks 1 and 2 were interconnected, EECOM had to constantly monitor the tanks to diagnose how much oxygen was lost at each point in time and how much was still left (but was at risk of also dissipating). Because of this imminent threat and the acuteness of the issue, FLIGHT decreased his communication frequency immediately by a little, giving room to increase EECOM's activity.

All members in charge reduced their communication quickly after the incident, focusing on trying to figure out what had happened and how severe the issue was. This involved monitoring systems and monitors. As a result, not only did FLIGHT reduce communication intensity, but CDR (the commander who leads the crew and announced the incident) did as well. There is an increase in *indegree sender*, suggesting that the individuals who had been the target of much past communication become more likely to send communication as well. This makes sense in this situation, as the individuals who had the most information about what was going on, would both be the target of requests for information and would be more likely to share their info with others in response. This is only needed for the first short time after the incident, after which this effect drops like we saw a drop in the other statistics as well. A further interesting finding is that the reduced communication of many of the parties also caused a drop in communication between Houston and the crew.

The next set of CPs is observable around $t = 0.42$ (58:04:50) for reciprocity, same location, LMP, TELMU, and FLIGHT. The quick draining of the oxygen levels in the tanks meant that it would become almost impossible to keep the Command and Service Module (CSM) operational. The solution mission control came up with was to use the lunar module as a "lifeboat" instead. Where EECOM was in charge of monitoring the CSM's condition, TELMU was EECOM's counterpart for the Lunar Module, watching over the LM's life-support and power systems just as EECOM did for the Command Module. TELMU played an important role in the implementation of the plan to use the Apollo 13 LM as a lifeboat. This set of CPs characterizes the effects of both teams' efforts to power down CSM as well as power the LM internally from its own power and using the LM as a lifeboat, a few minutes after TELMU informed FLIGHT about this plan. In order to get this procedure in place, two things needed to happen. First, mission control needed to come up with

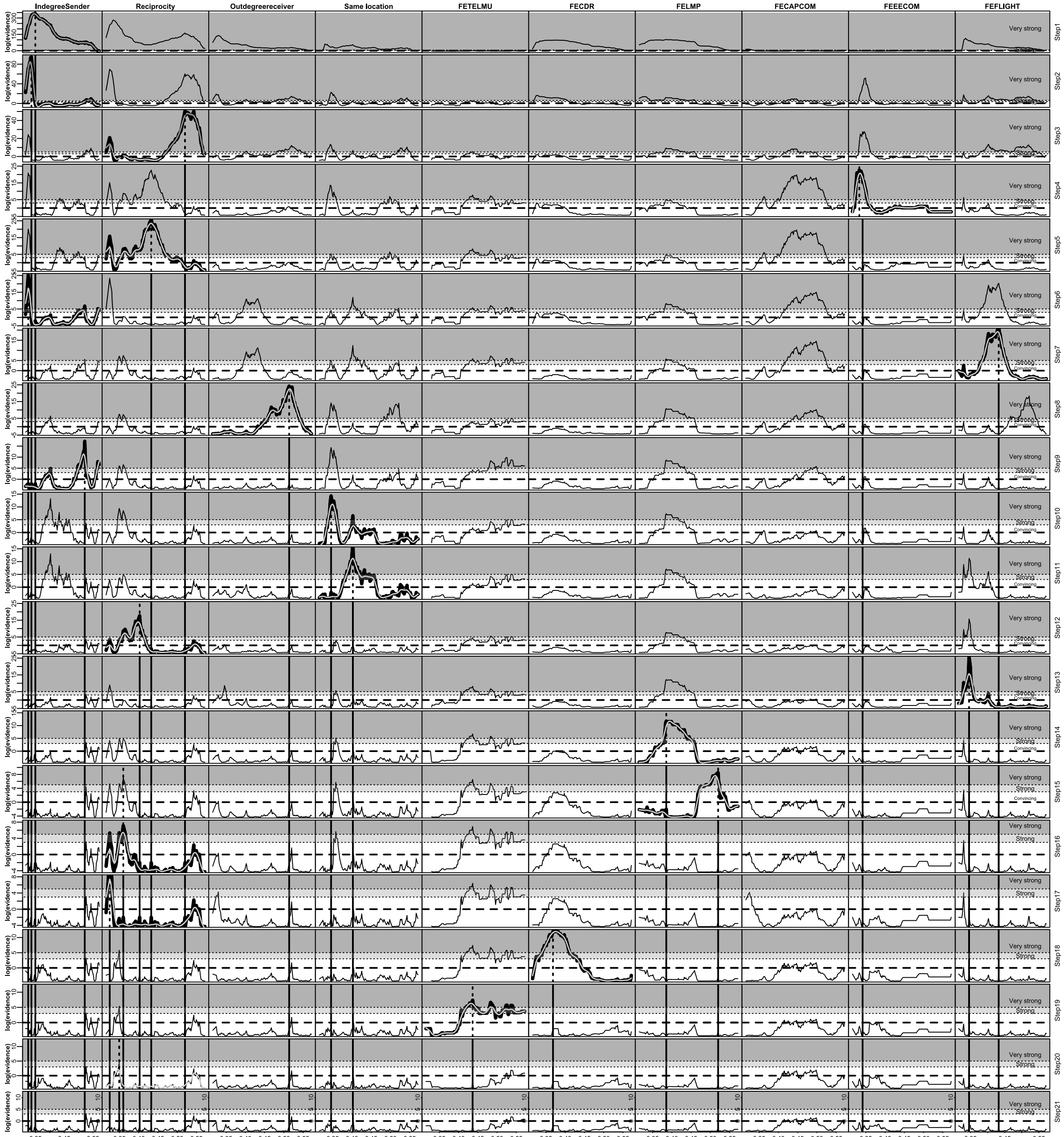


Fig. 7. CP evidence graph. The results of the CP detection algorithm for the Apollo 13 data. In all panels, the log-BF (y-axes) are plotted against the relational event times (x-axes). Each column contains a statistic in the model. The dark gray area corresponds to "very strong" evidence while the light gray and white area show the "strong" and "convincing" zones respectively. Each row corresponds to the appropriate step of the detection algorithm. The vertical dot-dashed lines refer to the time point with maximum evidence, while the vertical solid black lines indicate the detected CP in each step.

a good procedure and approach for the implementation. This shifted the communication to occur much more within-location than between locations; especially communication between the members of mission control intensified. Because the lunar module now became the center of attention (where it had played a less central role up to that point), the TELMU became more active in the communication and the flight director (FLIGHT) decreased his communication drastically. The latter is probably also caused by the new shift coming in (mission control operated in three shifts) and the flight director deciding to discuss part of the development of the rescue plan "offline" quietly. At the level of the crew in space, the commander (CDR) had already decreased the frequency of his communication and now the LMP (Fred Haise, who was the pilot of the lunar module that now needed to be made ready as a lifeboat) had to become much more involved in the communication.

The final set of CPs centers around $t = 0.52$ (60:28:50) for reciprocity, indegreesender, outdegrereciver, and LMP. The crew got informed about mission control's plan. Based on the plan, because they were, at that time, in 'water critical' situation in the LM, they decided to use as little water as possible. In order to do that they made a free-return maneuver of 16 ft per second at 61 h, which is almost 37 min from this time point. These are also visible in the data where at 60:22:58 CAPCAM was talking to CDR. Then they powered down Apollo's primary guidance, navigation, and control system (PGNCS, pronounced "pings"). Afterwards, they made another abort maneuver to powered down the PGNCS as soon as possible. All of this could be done with much less involvement by LMP than was needed in the previous few hours before this CP period, so this here we mainly see

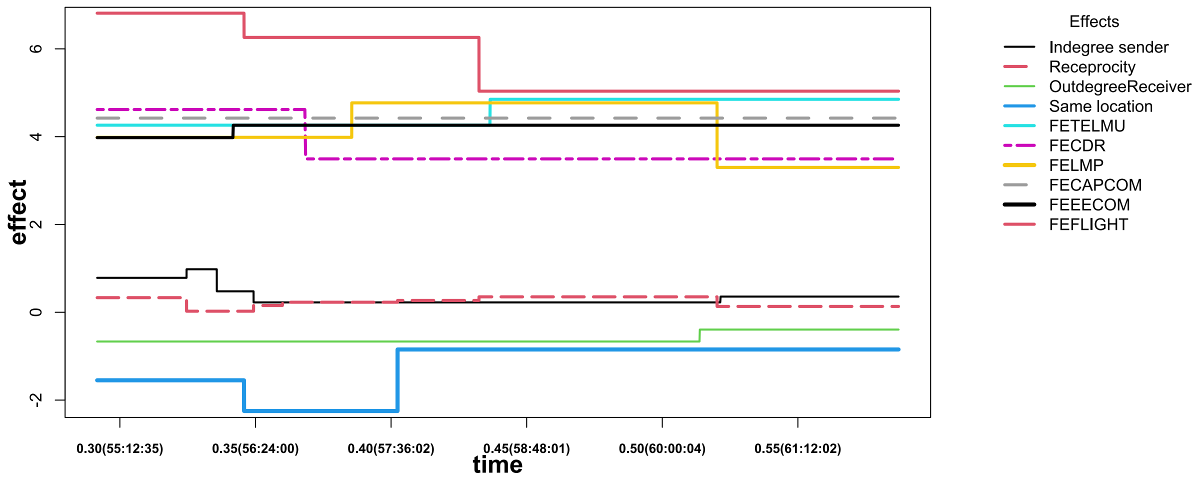


Fig. 8. Estimated parameters (effects) as a step function of time. In this plot y-axis shows the size of all effects over the entire event history time (x-axes). All CPs are located on the time points where the effects size jumps (change).

a sizeable decrease in LMP communication intensity. The other CPs at this time are fairly minor.

Overall, a much more refined analysis of the communication dynamics is possible by interpreting each CP individually and at its exact point in time. However, this is beyond the purpose of our illustration. Moreover, we contend that it is often possibly more useful to interpret groups of CPs (if they co-occur within a interpretably narrow time period) than to interpret each CP separately at its exact location. In part this is because the mathematical definition of the statistics causes them to respond somewhat faster/slower to nuanced changes in the event sequence (so they may not occur at the exact time anyway) and, in part, because the co-occurrence of CPs is informative and powerful in its own right.

Fig. 8 illustrates the dynamics of all effects based upon the detected CPs. The magnitude of each effect changes at each CP, which indicates that each effect is a type of step function over time. Overall, this plot depicts: (i) there are three main time zones, around $t = 0.33$ (55:55:10), $t = 0.42$ (58:04:50) and $t = 0.52$ (60:28:50) of the event history where interaction dynamics change explicitly, (ii) which effects show the largest/smallest change, (iii) the strengths of some effects (indegree sender, reciprocity, same location, and LMP), tend to go back, after the middle period, in the direction of the original state (from the beginning of the observation period), and (iv) some effects (such as outdegree receiver, reciprocity, and EECOM) have changes (jumps) at each CP, but the changes are only small. Observations (iii) and (iv) can be an indicator that the communication appears to veer back to (stay in) the original routine. When approaching to the end of the mission, it seems that most of the threatening problems were under control and the situation seemed less critical than before and had become more stable. Around the last change (last CP) the Apollo spacecraft control teams were in full swing making a plan on how the spacecraft can come back to Earth and have a safe landing and the initial scare had been overcome. FLIGHT announced all to refer to the checklist and thereafter listen to the loop and answer the related questions accordingly.

From Fig. 9 it becomes quite clear how the effects of the statistics change as step functions over time. The conventional REM (without any CPs) would assume a constant effect for each statistic over the entire observation period, the REM-CP results show that this is an overly strict assumption for these data.

The three CP time zones illustrate that there are four stable periods in the Apollo 13 data: the routine period before the first CP, and the periods between CP1 and CP2, between CP2 and CP3, and after CP3. We plot the accompanying aggregated networks of Apollo 13's voice loops data in Fig. 10. Edges (ties) display the message between actors (s, r) and the thickness of the ties is proportional to the number of messages exchanged between the actors. The figure clearly illustrates that the communication in Apollo 13's voice loops data is highly dynamic and has a distinct structure in the four periods.

5. Conclusions

An extension of the relational event model (REM-CP) was proposed which allows abrupt changes of the parameters using change points (CPs) in the observational period. To test whether a CP occurred at a given time point for a given statistic, a confirmatory Bayesian statistical test was proposed that can be used to quantify the relative evidence in the data for (non)existence of a CP at a time point for a specific network statistic. Note that frequentist significance tests cannot be used for quantifying the evidence for nonexistence because significance tests cannot distinguish between 'absence of evidence' and 'evidence of absence' of an effect (Dienes (2014)). Furthermore, an exploratory CP detection algorithm was proposed for finding CPs for all statistics over the entire observation period (a researcher could also limit the search to a specific part of the observation period if that would be of specific interest). The algorithm searches for the time point and statistic that results in the largest evidence to be a CP. If this largest evidence exceeds a prespecified threshold, the REM-CP is extended with an additional CP at this time point for this specific statistic, and the process is repeated until there is not enough evidence to extend the model with further CPs. Our analyses of synthetic data (considering three scenarios) showed that the detection algorithm can detect the actual CPs very well.

The confirmatory test, as well as the exploratory algorithm, were applied to the relational event history of communication messages between actors during NASA's famous, but troublesome, Apollo 13 mission, with the goal to better understanding how social interaction behavior and communication routines change during critical and noncritical situations. These analyses detected CPs that are consistent with what happened during the mission and can be explained by its highly dynamic and problematic situation. One thing *again* became quite clear after (or even during) Apollo 13's flight: space travel is far from routine. Nevertheless, our brief analysis of the data also showed that highly trained routines can allow teams to overcome some of the worst setbacks, such as being shipwrecked in space. Everyone involved in the Apollo program had been thoroughly trained in a common problem-solving method (known as Kepner-Tregeo training) that enabled those addressing highly complex and urgent problems to work together as effective ad hoc teams with little or no preparation (Francis and Tsekouras (2020)). Moreover, mission control and crew go through very extensive simulations before any mission, and many catastrophic problems are simulated, making the entire team involved

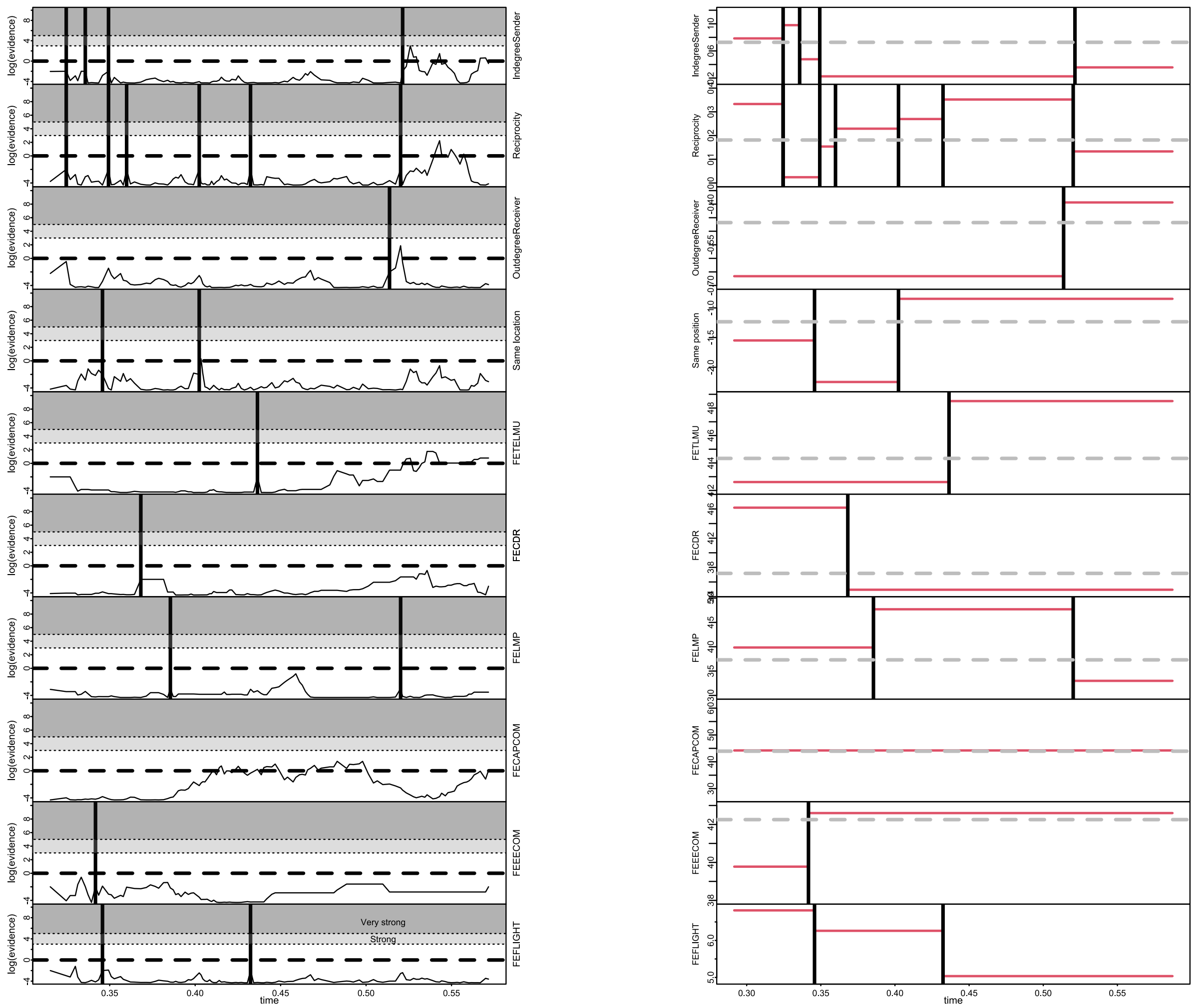


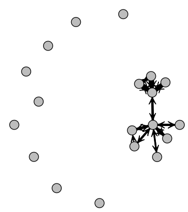
Fig. 9. Statistic specific change point(s). In the left column y-axes show the log evidence of the existence of a CP computed from Eq. (9) and the x-axes correspond to event times. In the right column, y-axes depict the size of all effects over the entire event history time (x-axes). In each row, the left panel illustrates the last step of our proposed algorithm for the specific statistic and the right panel indicates the corresponding effects as step function over time. The gray horizontal dashed line in the right column shows the corresponding effect estimated by conventional REM.

fully prepared for the worst. In fact, one of the simulations had actually involved the explosion of an oxygen tank in a very similar way to what ended up happening during the real mission and, although the crew had died (virtually) during the simulation, some of the learnings were later implemented during the real rescue mission. The Apollo mission was a very well-organized and well-prepared mission, with clear routines, clarity about decision ownership, extensive preparation, and highly skilled ground and space crew (Francis and Tsekouras (2020)). This also shows from the CP analysis: the CPs are all borne out of an organized and skilled approach to solving the life-threatening situation. A much less trained and less well-prepared team might have shown many more CPs with a much less clearly interpretable pattern.

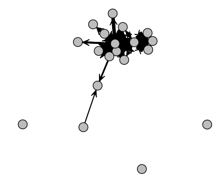
Moreover, the empirical application showed that the REM-CP methodology is a useful tool for segmenting relational event history data (based on the CPs) and modeling the data within each segment. In the analysis of the Apollo 13 data, this resulted in four stable periods over the observational period. This segmentation of the observational period allows for a deeper understanding of dynamic communication patterns among actors in temporal social networks and for subsequently analyzing (and comparing) the interaction patterns per stable segment.

Throughout this paper, the CP detection algorithm was applied using a threshold that corresponds to "very strong" evidence for the existence of a CP, according to the guidelines of Kass and Raftery (1995), which suggest a BF of at least 150. This implies that a CP is added when a model with this additional CP is at least 150 times more likely to have generated the data than a model without this additional CP. This threshold is relatively conservative which will ensure that the model is only extended in the case of very convincing evidence for an additional CP. Researchers could use a more liberal threshold value if preferred, such as a BF of 50 or even 20. This will result in a REM-CP with more CPs with potentially smaller "jumps", and thus in a less parsimonious model. This more liberal setup of the algorithm may be preferred if many smaller steps are anticipated for the model parameters over the observational period. Moreover, this setup could be considered to be a middle ground between modeling gradual changes in a REM using a moving window technique Mulder and Leenders (2019), Meijerink-Bosman et al. (2022) and abrupt changes as in the case of CPs with a conservative threshold as done in the current paper.

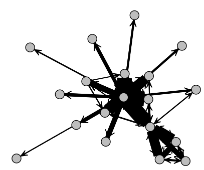
The main computational burden of the CP detection algorithm lies in fitting the model many times for all possible time points on the grid and for all possible network effects. Thus, if a fine (rough) grid is used, the computational costs are relatively high (low), and if a model is considered with many (few) network effects, the computational costs are relatively high (low). Note that the separate tasks, i.e., computing the evidence for the existence of a CP, can be run concurrently, allowing the researcher to take advantage of parallel processing. When using a modern computer cluster, the computational cost for our real Apollo 13 data was relatively low, taking approximately 2 min to run the model



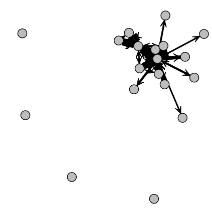
(a) Aggregated networks of Apollo data: first segment



(b) Aggregated networks of Apollo data: second segment .



(c) Aggregated networks of Apollo data: third segment.



(d) Aggregated networks of Apollo data: Fourth segment.

Fig. 10. Aggregated networks of Apollo 13's voice loops data over (a) the first inferred segment, (b) the second inferred segment, (c) the third inferred segment, and (d) the fourth segment. The learned CPs are around $t = 0.33$ (55:55:10), $t = 0.42$ (58:04:50) and $t = 0.52$ (60:28:50).

at a given time point on a Blade Server with R version 4.2.2. Further note that the computational time to fit one model, and thus to compute the evidence for the existence of a CP for one specific parameter at one specific time point depends on the estimation algorithm that is used (here we used the algorithm implemented in the *relevent* package). If this single step is computationally expensive (due to a large network, a large event sequence, or a large model), one could start with a relatively rough grid to search for CPs which would be computationally feasible. Finally, note that the burden to compute the evidence (using the BF based on Gaussian approximations, Gu et al. (2018) is negligible which was also the motivation for this choice.

In the current paper, we allowed CPs to only occur at the observed event times to abide the piecewise constant hazard property of the model where the event rates can only change at the observed events. It is possible however to drop this assumption, which may be reasonable in certain real-life examples. To model this, we would need to split the waiting time before and after the CP where the first part is modeled as a waiting time where no events were observed, and the waiting time for the second part starts at the CP and ends at the first event after the CP. The proposed methodology for testing and detecting these inter-event CPs can then be executed in a similar fashion as discussed.

It is also relevant to note that the proposed methodology for testing and detecting CPs does not take the content of the messages into account. The content, however, might have additional unique information that could also inform us about the presence of CPs that do not cause changes in the timing and the dyads in the observed sequence. For example, certain messages have a clear positive sentiment (e.g., "Houston. Trajectory is good; thrust is good".) whereas other messages have a clear negative sentiment ("Houston, we've had a problem here".). Extending the proposed methodology to also include the content of relational events to test and detect CPs would therefore be an interesting direction for future research. Other interesting extensions include the detection of change points of memory decay functions in relational event models (Arena et al. (2023), Arena et al. (2022)), e.g., to identify abrupt changes about the weighting of past events in the endogenous statistics as a result of critical situations, or combining the methodology with regularization techniques for relational event models (Karimova et al. (2023)), to obtain more parsimonious REM-CPs. An area for future research could also involve the utilization of the Bayesian multiple change point process, in conjunction with the Reversible-Jump Markov Chain Monte Carlo algorithm (Shafiee Kamalabad and Grzegorczyk (2021)), to derive the posterior probabilities of change points and infer them from the data.

To ensure the usage of the new algorithm to better understand complex dynamic social interaction behavior in other real-life applications, the methodology will be implemented in a new R package, called *'remverse,'* which is scheduled for later this year. This package will, among other things, allow social network researchers to study dynamic social interaction behavior by extending REMs with CPs in a principled probabilistic manner. Note that the proposed methodology is rather generic and can be embedded into the other models as well and is thus not only limited to the REMs.

The proposed model is particularly useful for applications where social network dynamics abruptly changes over time. The abrupt changes in social networks could result from, for example, critical incidents occurring during a space mission, an unexpected physical response by a patient during surgery, an intervention in an organization, the start of lunchtime that causes employees to cease their regular work and spend some off-time at the cafeteria, a teacher announcing an unexpected test to her pupils in the classroom, emergency responders being active at an emergency site, et cetera. When CPs occur, models that assume a stability of the effects will provide biased results and under-inform the researcher as to what is happening in the data. The methodology that is proposed in this paper can be used to detect CPs or test for specific CPs, and thereby extend the regular REM to a REM-CP is needed. In addition, it can often be useful to run the CP detection algorithm as an exploratory analysis, to check if the foundational assumption of effect stability (Leenders et al. (2016)) can be maintained. If the algorithm suggests no relevant CPs, the regular REM is sufficient and applicable. If there are meaningful CPs, however, this tells the researcher that the conventional REM is not appropriate and the REM-CP should be used for a more fine-grained analysis of social interaction dynamics in the network.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to thank the review team for their feedback which resulted in important improvements of the manuscript. This work was supported by an ERC Starting Grant (758791).

References

Apollo 13 Real-time, (n.d.). http://apollo13realtime.org/.

Apollo Flight Journal, 2015. https://history.nasa.gov/afj/ap13fj/07day3-before-the-storm.html.

Arena, G., Mulder, J., Leenders, R.T.A., 2022. A bayesian semi-parametric approach for modeling memory decay in dynamic social networks. Sociol. Methods Res. 00491241221113875.

Arena, G., Mulder, J., Leenders, R., 2023. How fast do we forget our past social interactions? Understanding memory retention with parametric decays in relational event models. Netw. Sci. (in press).

Butts, C.T., 2008. 4. A relational event framework for social action. Sociol. Methodol. 38 (1), 155–200.

Chapman, M.T., Lines, R.L., Crane, M., Ducker, K.J., Ntoumanis, N., Peeling, P., Parker, S.K., Quested, E., Temby, P., Thøgersen-Ntoumani, C., et al., 2020. Team resilience: A scoping review of conceptual and empirical work. Work Stress 34 (1), 57–81.

Chen, X., Baron, M., et al., 2014. Change-point analysis of survival data with application in clinical trials. Open J. Statist. 4 (09), 663.

Dickey, J.M., 1971. The weighted likelihood ratio, linear hypotheses on normal location parameters. Ann. Math. Stat. 204–223.

Dienes, Z., 2014. Using Bayes to get the most out of non-significant results. Front. Psychol. 5, 781.

Francis, D., Tsekouras, G., 2020. Apollo 13 - Crisis, innovation and sensemaking. In: The ISPIM Innovation Conference – Innovating in Times of Crisis. ISBN: 9789523354661, Retrieved from https://www.ispim-virtual.com, XXXI ISPIM Innovation Conference : Innovating in times of Crisis, XXXI ISPIM ; Conference date: 07-06-2020 Through 10-06-2020.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4), 711–732.

Gu, X., Mulder, J., Hoijtink, H., 2018. Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. Br. J. Math. Stat. Psychol. 71 (2), 229–261.

Hartwig, A., Clarke, S., Johnson, S., Willis, S., 2020. Workplace team resilience: A systematic review and conceptual development. Organ. Psychol. Rev. 10 (3–4), 169–200.

Hinkley, D.V., 1970. Inference about the Change-Point in a Sequence of Random Variables. Oxford University Press.

Janssens, M., Meslec, N., Leenders, R.T.A., 2022. Collective intelligence in teams: Contextualizing collective intelligent behavior over time. Front. Psychol. 13, 989572.

Jeffreys, H., 1961. Theory of Probability, third ed. Oxford, Clarendon.

Karimova, D., Leenders, R., Meijerink-Bosman, M., Mulder, J., 2023. Separating the wheat from the chaff: Bayesian regularization in dynamic social networks. Netw. Sci. (in press).

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90 (430), 773–795.

Kratzer, J., Leenders, R.T.A., Van Engelen, J.M., 2006. Managing creative team performance in virtual environments: an empirical study in 44 R&D teams. Technovation 26 (1), 42–49.

Larson, L., Wojcik, H., Gokhman, I., DeChurch, L., Bell, S., Contractor, N., 2019. Team performance in space crews: Houston, we have a teamwork problem. Acta Astronaut. 161, 108–114.

Leenders, R.T.A., Contractor, N.S., DeChurch, L.A., 2016. Once upon a time: Understanding team processes as relational event networks. Organ. Psychol. Rev. 6 (1), 92–115.

Liang, K.-Y., Self, S.G., Liu, X., 1990. The cox proportional hazards model with change point: An epidemiologic application. Biometrics 783–793.

Luo, X., Turnbull, B.W., Clark, L.C., 1997. Likelihood ratio tests for a changepoint with survival data. Biometrika 84 (3), 555–565.

Meijerink-Bosman, M., Leenders, R., Mulder, J., 2022. Dynamic relational event modeling: Testing, exploring, and applying. PLoS One 17 (8), e0272309.

Mulder, J., 2014. Prior adjusted default Bayes factors for testing (in) equality constrained hypotheses. Comput. Statist. Data Anal. 71, 448–463.

Mulder, J., Leenders, R.T., 2019. Modeling the evolution of interaction behavior in social networks: A dynamic relational event approach for real-time analysis. Chaos Solitons Fractals Nonlinear 119, 73–85. http://dx.doi.org/10.1016/j.chaos.2018.11.027.

O'Hagan, A., 1995. Fractional Bayes factors for model comparison. J. R. Stat. Soc. Ser. B Stat. Methodol. 57 (1), 99–118.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociol. Methodol. 111–163.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 461–464.

Shafiee Kamalabad, M., Grzegorczyk, M., 2020. Non-homogeneous dynamic Bayesian networks with edge-wise sequentially coupled parameters. Bioinformatics 36 (4), 1198–1207.

Shafiee Kamalabad, M., Grzegorczyk, M., 2021. A new Bayesian piecewise linear regression model for dynamic network reconstruction. BMC Bioinformatics 22 (2), 1–24.

Shiryaev, A.N., 1963. On optimum methods in quickest detection problems. Theory Probab. Appl. 8 (1), 22–46.

Tseng, I., 2017. Apollo 13 Real-Time. GitHub, https://github.com/issa-tseng/apollo13rt.

van den Oever, F., Schraagen, J.M., 2021. Team communication patterns in critical situations. J. Cogn. Eng. Decis. Mak. 1555343420986657.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N.J., Gronau, Q.F., Haaf, J.M., et al., 2021. The JASP guidelines for conducting and reporting a Bayesian analysis. Psychon. Bull. Rev. 28 (3), 813–826.

Wang, B., Wang, X., Song, L., 2019. Estimation in the single change-point hazard function for interval-censored data with a cure fraction. J. Appl. Stat..

Wetzels, R., Grasman, R.P., Wagenmakers, E.-J., 2010. An encompassing prior generalization of the Savage–Dickey density ratio. Comput. Statist. Data Anal. 54 (9), 2094–2102.

Wikipedia contributors, 2022. Apollo 13 — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Apollo_13&oldid=1070171476 [Online; accessed 10-February-2022].

Woods, D.D., 2006. Essential characteristics of resilience for organizations. In: Resilience Engineering: Concepts and Precepts. Ashgate, Aldershot, UK, pp. 21–34.

Woods, D.D., 2015. Four concepts for resilience and the implications for the future of resilience engineering. Reliab. Eng. Syst. Saf. 141, 5–9.

Woods, D.D., Branlat, M., 2011. Basic patterns in how adaptive systems fail. In: Resilience Engineering in Practice. Ashgate, Farnham, UK, pp. 127–144.