

# NETWORKS 2021:

A JOINT SUNBELT AND NETSCI CONFERENCE

JULY 6-11, 2021 | WASHINGTON D.C.

OpenConf Peer Review and Submission Management System

[OpenConf Home](#) [Privacy Policy](#) [Email Chair](#)

## Submission

**Thank you for your submission. Your submission ID number is 10417. Please write this number down and include it in any communications with us.**

**Below is the information submitted. We have also emailed a copy to the submission contact. If you notice any problems or do *not* receive the email within 24 hours, please contact us.**

Submission ID: 10417

Title: Triads, Dyads, and Gene Functions: When Social Network Analysis meets Phylogenetics

Author 1:

First Name: George

Last Name: Vega Yon

Department/Division: Department of Preventive Medicine

Organization: University of Southern California

Country: United States

Email: vegayon@usc.edu

Presenter: Yes

Author 2:

First Name: Duncan

Last Name: Thomas

Department/Division: Department of Preventive Medicine

Organization: University of Southern California

Country: United States

Email: dthomas@usc.edu

Presenter: No

Author 3:

First Name: Huaiyu

Last Name: Mi

Department/Division: Department of Preventive Medicine

Organization: University of Southern California

Country: United States

Email: huaiyumi@usc.edu

Presenter: No

Author 4:

First Name: John

Last Name: Morrison

Department/Division: Department of Preventive Medicine

Organization: University of Southern California

Country: United States

Email: jmorrr@usc.edu

Presenter: No

Author 5:

First Name: Paul  
Last Name: Thomas  
Department/Division: Department of Preventive Medicine  
Organization: University of Southern California  
Country: United States  
Email: pdthomas@usc.edu  
Presenter: No

Author 6:  
First Name: Paul  
Last Name: Marjoram  
Department/Division: Department of Preventive Medicine  
Organization: University of Southern California  
Country: United States  
Email: pmarjora@usc.edu  
Presenter: No

Contact Author: Author 1

Topic(s):  
- Biological Networks  
- Complex Systems  
- Dynamics  
- Statistical Advances

Organized:

Type: Oral Presentation (15 mins)

Poster Option: Yes, I am willing to be considered for a poster session.

Keywords: statistical methods, exponential random graphs, complex systems, gene functions, evolution

Abstract: Understanding what genes do is undoubtedly one of the most pressing challenges of biomedical research. Knowing gene functions is paramount for advancing treatment development, yet, conducting experiments to augment our knowledge is both time-consuming and, many times, implausible. In this context, scientists are in a race to automatically annotate gene functions at a large scale. This paper presents a novel approach that taps into phylogenetic trees and methods developed in social network analysis to build an evolutionary model of gene functions.

In simple terms, phylogenetic-based prediction models allow us to borrow known gene annotations to infer others without the need to run experiments. Intuitively, if gene A's evolutionary cousin has function X, the chances are that A has the function too. Using a Markov transition model, we can characterize the evolutionary process at each step, i.e., when functions are gained or lost throughout evolution. Notwithstanding, as the number of functions we look at increases--genes can have multiple functions,--Markov transition models rapidly fall to the curse of dimensionality. Instead, inspired by the ideas developed around Exponential Random Graph Models (ERGMs), we propose using sufficient statistics to simplify the problem while improving its interpretability significantly.

Without sufficient statistics, a common approach to fit this model would require estimating  $2^{(3k)}$  parameters in a problem with k gene functions. For example, if we were looking at four different functions, the procedure would involve 4096 parameters. Using sufficient statistics, we can reduce the number of parameters to as little as 15 and still incorporate complex interplay between gene functions. The only thing we ought to do is treat the transition process the same way we would do a social network. We will explain the details in our presentation.

In summary, we propose a novel approach for modeling gene function evolution using phylogenetic trees and ERGMs. This approach's main contribution is two three-fold: It bypasses the curse of dimensionality experienced in canonical approaches, it provides a way to conduct hypothesis testing of different theories of evolution, and most importantly, it has the prospect of enhancing our prediction capacity using an interpretable model.

Previous:  
- I have attended a Sunbelt meeting before.  
- I have attended a NetSci meeting before.

Present: I anticipate this presentation will be delivered remotely through an online version.

Certainty: Somewhat certain of my answer above

Comments:

Consent: I consent to the collection and use of my personal information, including receiving emails, consistent with the Privacy Policy linked above. I have also obtained the consent of all other individuals whose information I provide.

File: uploaded