

## George G. Vega Yon, Ph.D.

mo: (626) 381-8171 | email: g.vegayon@gmail.com | Sandy, UT, 84094

<https://ggv.cl/> / <https://github.com/gvegayon> / <https://linkedin.com/in/georgevegayon> / <https://ggv.cl/scholar>

### PROFESSIONAL SUMMARY

Accomplished **data scientist** with over **10 years of experience** with multiple software packages and scientific publications. A results-driven professional with over **half a million downloads** and over **200 citations for academic work**. A solution development and delivery champion with a proven track record of spearheading solutions across data science, network science, machine learning, and statistics.

### EXPERIENCE

#### **Assistant Professor of Data Science, The University of Utah**

**Nov 2021 – Present**

- Studies Machine Learning, Network Science, Data Science, Statistical Computing.
- Managed a team of researchers, leading to published papers, software, and conference talks.
- Taught and designed the first course on HPC using R and C++ (graduate level).
- Founder of the "Network Science and Social Network Analysis at the U" (NetSNAU) Group.
- Contributed to research grants (CDC and VA,) helping to secure over 1 MM USD in funding.
- A core faculty member of the "Utah Center for Data Science."

#### **Research Programmer II, University of Southern California**

**Feb 2018 – Nov 2021**

- Provides technical support and training sessions on software development, HPC, R, and C++.
- Write scientific papers on network science, machine learning, and statistics and present them at conferences.
- Designed and taught the course "Intro to Health Data Science" (graduate level).
- Contributed to research grants (NIH and DoD,) helping to secure over 10 MM USD in funding.

#### **Programmer Analyst II, University of Southern California**

**Oct 2015 – Feb 2018**

- Served as a consultant in statistics and computer science across campus.
- Founder of the "R Bootcamp for Statistical Computing."
- Wrote scientific papers and software on network science and presented them at conferences.
- Designed and led workshops on R and Social Network Analysis.

#### **Graduate Student Researcher, California Institute of Technology**

**Aug 2014 – Oct 2015**

- Data Scientist studying social network activity of US congress.
- Created R package using Twitter API and a web scraper tool for extracting twitter accounts from websites.

#### **Research Analyst, Chilean Pension Supervisor**

**Aug 2011 – Aug 2014**

- Wrote papers and automatized statistical reports about the Chilean unemployment insurance system.
- Managed social security records and created representative samples for researchers.
- Designed and implemented a pipeline for simulation and forecasting of the unemployment insurance government funds. Reports were distributed to the Chilean Congress.

### EDUCATION

**Ph.D. in Biostatistics**, University of Southern California, 2020

**M.Sc. in Economics**, California Institute of Technology, 2016

**MA in Economics and Public Policy**, Universidad Adolfo Ibáñez, 2011

**BS in Business Administration**, Universidad Adolfo Ibáñez, 2011

### SKILLS

R, C++, LaTeX, SQL, Python, XML, NLP, Stata, AWS, Git, GitHub, Docker, TensorFlow, PyTorch, Deep Learning, continuous integration, Slurm, Unix, Jira, Scrum, Kanban, team management and coordination, R Shiny, excellent communication skills.

## SOFTWARE PACKAGES (selected)

- *epiworldR: Fast Agent-Based Epi Models* (2023). R package version 0.2-1 URL: <https://cran.r-project.org/package=epiworldR> .
- *rgexf: Build, Import and Export GEXF Graph Files* (2020). R package version 0.16.0. URL: <https://CRAN.R-project.org/package=rgexf>.
- *netdiffuseR: Analysis of Diffusion and Contagion Processes on Networks* (2020). R package version 1.22.0. URL: <https://cran.r-project.org/package=netdiffuseR>.
- *ergmito: Exponential Random Graph Models for Small Networks* (2020). R package version 0.3-0. URL: <https://cran.r-project.org/package=ergmito>.
- *slurmR: A Lightweight Wrapper for 'Slurm'* (2020). R package version 0.4-1. URL: <https://CRAN.R-project.org/package=slurmR>.
- *fmcmc: A friendly MCMC framework* (2020). R package version 0.3-0. URL: <https://CRAN.R-project.org/package=fmcmc> .

## ACADEMIC PUBLICATIONS (selected)

- George G. Vega Yon. “Power and Multicollinearity in Small Networks: A Discussion of “Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks” by Krivitsky, Coletti & Hens.” In: **Journal of The American Statistical Association** (2023). to appear.
- George G. Vega Yon, Mary Jo Pugh, and Thomas W. Valente. *Discrete Exponential-Family Models for Multivariate Binary Outcomes*. Nov. 2022. In: **arXiv**: [2211.00627 \[cs, stat\]](https://arxiv.org/abs/2211.00627).
- George G. Vega Yon, Andrew Slaughter, and Kayla de la Haye. “Exponential random graph models for little networks.” In: **Social Networks** 64 (2021), pp. 225–238. URL: <https://doi.org/10.1016/j.socnet.2020.07.005>.
- George G. Vega Yon, Duncan C. Thomas, John Morrison, Huaiyu Mi, et al. “Bayesian parameter estimation for automatic annotation of gene functions using observational data and phylogenetic trees.” In: **PLOS Computational Biology** 17.2 (Feb. 2021), pp. 1–35. URL: <https://doi.org/10.1371/journal.pcbi.1007948>.
- George G. Vega Yon and Paul Marjoram. “*fmcmc: A friendly MCMC framework*.” In: **Journal of Open Source Software** 4.39 (July 2019), p. 1427. URL: <http://joss.theoj.org/papers/10.21105/joss.01427>.
- George G. Vega Yon and Brian Quistorff. “*parallel: A command for parallel computing*.” In: **The Stata Journal: Promoting communications on statistics and Stata** 19.3 (Sept. 2019), pp. 667–684. URL: <http://journals.sagepub.com/doi/10.1177/1536867X19874242>.

## AWARDS

- Best Paper Awards (Computational Methods,) 72 ICA conference, 2022.
- Travel Grant, Society of Young Network Scientist, 2019.
- Fellowship, California Institute of Technology, 2014.
- Honorable mention (posters session,) Sociedad de Economía de Chile, 2012.
- Scholarship, Adolfo Ibáñez University, 2006.

## REFEREE (ad hoc reviewer)

- Journal of The American Statistical Association
- BMC Infectious Diseases
- The Official Journal of The Society for Computational Economics
- The R Journal
- Social Networks
- Journal of Mathematical Sociology
- Journal of Open Source Software
- Bioinformatics
- Computer Methods and Programs in Biomedicine Update