SAM F. L. WINDELS

Postdoctoral Researcher in Network Biology and Data Integration

sam.windels@gmail.com +32 478 64 34 36 github.com/samwindels orcid.org/0000-0001-8443-5041

SUMMARY

- I am a Postdoctoral Researcher at the Barcelona Supercomputing Center in Nataša Pržulj's Integrated Network Biology group. Renowned for pioneering network analysis and data integration to extract biomedical insights from large-scale molecular and clinical datasets, she has been awarded three prestigious ERC grants in the past decade, among other accolades. I completed my PhD in her group at University College London in 2021, which was ranked the 8th best university globally by the QS World University Rankings at the time.
- I have developed novel graphlet-based methods [J–1, J–4], and interpretable data integration methods [J–2, J–3] to extract biological insights from molecular networks, that have been published in top-tier journals such as Oxford Bioinformatics (IF 4.4 in 2023) [J–1, J–3], PLOS One (IF 2.3 in 2023) [J–4] and Nature Communications (IF 14.7 in 2023) [J–2]. I have also contributed to a book chapter on machine learning for data integration in cancer precision medicine [BC–1]. My research has been cited 115 times and I have an h-index of 5.
- My current research focuses on building more powerful graphlet-based network descriptors to create more powerful yet descriptive methods for network analysis. Recently, I introduced the concept of graphlet orbit adjacency to characterize the topology (i.e., structure and wiring) implicitly captured by random walks and to examine their limitations in understanding topology-function relationships in networks [WP–2]. Building on this work, I am now leveraging orbit adjacency to analyze the topology captured by Graph Neural Networks (GNNs) and to identify their limitations in learning topological features.
- Over the years, I have supervised four PhD students on projects involving drug repurposing, functional annotation of embedding spaces, text mining of DNA sequences and the integration of multi-omics data.

EDUCATION

PhD in Computational Network Biology — University College London — 2016 – 2021

- Thesis: Graphlet-adjacency provides complementary views on the functional organisation of the cell and cancer mechanisms.
- Supervisor: Nataša Pržulj
- Received a Computer Science Studentship from the Department of Computer Science, UCL.

MSc in Computing Science — Imperial College London — 2014 – 2015

- Thesis: Signalling pathway reconstruction by data fusion.
- Focus on foundations of Computer Science and Machine Learning.

MSc in Business Engineering — Ghent University — 2012 – 2014

- Thesis: The resource constrained stochastic multi-project scheduling problem.
- Focus: Operations Research and Supply Chain Optimisation.
- Awards: Prof. Dr. Beyaert prize for most meritorious bursary student.
- Graduated Magna Cum Laude.

BSc in Business Engineering — Ghent University — 2008 – 2012

- Included an exchange semester at the Hochschule für Wirtschaft und Recht, Berlin.
- Graduated Cum Laude.

POSITIONS

Postdoctoral Researcher	Barcelona Supercomputing Center	2021 – now
PhD student	University College London	2016 – 2020

GRANTS AND AWARDS

JDC2023-004889-C	72,000 €	2025 – 2027

- Two-year post-doctoral fellowship for young researchers.
- Granted by the Spanish Ministry of Science and Innovation (MCIU/AEI/10.13039/501100011033).

Dr. Beyaert Prize	2,500 €	2016

- Prize for the most meritorious bursary student awarded by the Faculty of Economics and Business Administration, Ghent University.

PUBLICATIONS

Google Scholar Statistics (2nd of December 2024)

- Citations: 115
- h-index: 5
- i10-index: 3

Journal articles

[J–5] **Windels, Sam F.L.**, Daniel Tello Velasco, Mikhail Rotkevich, Noël Malod-Dognin, and Nataša Pržulj. "Graphlet-based hyperbolic embeddings capture evolutionary dynamics in genetic networks". In: *Bioinformatics* (2024), btae650.

[J–4] Sam FL Windels, Noël Malod-Dognin, and Nataša Pržulj. "Graphlet eigencentralities capture novel central roles of genes in pathways". In: *Plos one* 17.1 (2022), e0261676.

[J–3] **Windels, Sam F.L.**, Noël Malod-Dognin, and Nataša Pržulj. "Identifying cellular cancer mechanisms through pathway-driven data integration". In: *Bioinformatics* 38.18 (2022), pp. 4344–4351.

[J–2] Noël Malod-Dognin, Julia Petschnigg, **Windels, Sam F.L.**, Janez Povh, Harry Hemingway, Robin Ketteler, and Nataša Pržulj. "Towards a data-integrated cell". In: *Nature Communications* 10.1 (2019), pp. 1–13.

[J–1] **Windels, Sam F.L.**, Noël Malod-Dognin, and Nataša Pržulj. "Graphlet Laplacians for topology-function and topology-disease relationships". In: *Bioinformatics* 35.24 (2019), pp. 5226–5234.

Book chapters

[BC–1] Noël Malod-Dognin, **Windels, Sam F.L.**, and Nataša Pržulj. "Machine Learning for Data Integration in Cancer Precision Medicine: Matrix Factorization Approaches". In: *Analyzing Network Data in Biology and Medicine: An Interdisciplinary Textbook for Biological, Medical and Computational Scientists*. Cambridge University Press, 2019, pp. 286–312.

Working papers

[WP–3] Zuqi Li, **Windels, Sam F.L.**, Noel Malod-Dognin, Seth M Weinberg, Mary L Marazita, Susan Walsh, Mark D Shriver, David W Fardo, Peter Claes, Natasa Przulj, et al. "Clustering individuals using INMTD: a novel versatile multi-view embedding framework integrating omics and imaging data". Preprint: `https://www.biorxiv.org/content/10.1101/2024.09.23.614478v1`. 2024.

[WP–2] **Windels, Sam F.L.**, Noel Malod-Dognin, and Natasa Przulj. “Graphlets correct for the topological information missed by random walks”. Preprint: `https://arxiv.org/abs/2405.14194`. 2024.

[WP–1] Carme Zambrana, **Windels, Sam F.L.**, Noël Malod-Dognin, and Nataša Pržulj. “Between viral targets and differentially expressed genes in viral infections: the sweet spot of disease mechanisms for therapeutic intervention”. 2024.

SOFTWARE

GRADCO

github.com/samwindels/gradco 2024

- GRaphlet ADjacency Counter (GRADCO) is a python package for computing orbit adjacency matrices.
- GRADCO generalizes state-of-the-art graphlet counting and inferencing algorithms, implemented in C++, to efficiently process medium-sized networks (≈ 20,000 nodes) in less than than 5 minutes [WP–2].
- GRADCO can output all graphlet-based node-level statistics for undirected networks.
- GRADCO was installed 1,026 times during November 2024, see PyPiStats.

ENRICO

github.com/samwindels/enrico 2024

- ENRIchment COmputer (ENRICO) is vectorized cluster enrichment analysis framework available in Python.
- ENRICO scales to clusterings of millions of data points with thousands of annotations [WP–3].

PHD STUDENT SUPERVISION

Carme Zambrana

Universitat Politècnica de Catalunya 2021 – 2024

- Topic: new methods for mining heterogeneous omics data for drug repurposing in viral diseases.
- Graduated with honours in 2024.
- Currently a postdoctoral researcher at the ICONBI research group at the Barcelona Supercomputing Centre.

Mikhail Rotkevich

Universitat Politècnica de Catalunya 2021 – 2024

- Topic: Applying NLP based methods to the DNA within the context of precision medicine.
- Graduated in 2024.
- Currently a postdoctoral researcher at the Centre for Genomic Regulation (CRG), Barcelona.

Daniel Tello

Universitat Politècnica de Catalunya 2021 – 2023

- Topic: Developing new methods to functionally annotate network embeddings.
- Graduated with honours in 2023.
- Associate researcher in the Genomics and Therapeutics group at the FIBH120.

Zuqi Li

K.U. Leuven, Belgium 2023

- Zuqi, visited our group for a three-month secondment via the TranSYS exchange program, part of the Horizon 2020 framework, grant 860895.
- Topic: Integrating multi-omics and imaging data for metagenomics.

TEACHING

Assistant Advanced Deep Learning

University College London 2^{Nd} term 2018

- Course on supervised learning with neural networks with applications in object recognition and NLP.
- I was responsible for marking coursework and exams.

Assistant Computer Vision — University College London — 1st term 2018

- Course on computational methods for object tracking, object recognition and 3D modelling from images.
- I assisted students in understanding and solving the weekly courseworks.

Assistant Web Economics — University College London — 2nd term 2017

- Course on computational methods for online auctioning, user behavior targeting and social media mining.
- I was co-responsible for the design, supervision and marking of the coursework.

Code Instructor — Turinglab — 2016 – 2019

- Turinglab is an NGO that teaches Javascript and Python to children aged 10 to 14.

CONFERENCES

Conference Workshop BelBi — Belgrade, Serbia — 2024

- At the Belgrade Bioinformatics (BelBi) conference, we organised a workshop on multi-omics data integration for the identification of disease mechanisms using graphlet-based topological signatures.
- I designed and implemented the workshop's materials on topological network analysis.

WORK EXPERIENCE

Junior Data Scientist — Conundra — 2015 – 2016

- Conundra is a Ghent based tech-startup that solves routing optimization problems.
- I built a routing simulator to estimate the road-tax fee based on GPS tracking data.
- I implemented a business intelligence platform to offer interactive access to different views of customer data.

SKILLS

High Performance Computing

- Experienced with **SnakeMake** for creating reproducible workflows and submitting jobs to the MareNostrum supercomputer using **SLURM**.

Command Line Tools

- Proficient in managing Python environments with **MicroMamba** (similar to Conda).
- Daily user of **Vim**, **TMUX**, and **iPython** for development and productivity.

Programming Languages

- Proficient: Python, C++, LaTeX
- Familiar: Java, Matlab, SQL

Languages

- Native: Dutch
- Near-Native: English (*IELTS 8/10*, 2013)
- Conversational: French, German