

## Bachelor- / Masterarbeit in Bioinformatik

Prof. Dr. Dominik Heider  
dominik.heider@uni-marburg.de

### Thema: Maschinelle Lernverfahren zur Analyse und Identifikation von spezi- fischen Netzwerken

In natürlich vorkommenden Ökosysteme interagieren Organismen verschiedener Arten miteinander und mit ihrer unbelebten Umwelt. Das bedeutet, dass sich nicht nur die Organismen untereinander beeinflussen (z.B. durch Prädation oder Symbiose), sondern die Lebensfähigkeit eines Organismus auch von dem Zustand der Umwelt abhängt und diese Umwelt wiederum von den Organismen geformt wird [1]. Werden Ökosysteme auf diese Weise betrachtet, dann leuchtet es ein, dass der Zustand des Interaktionsnetzwerkes eines Ökosystems ein starker Indikator für den Gesundheitszustand des betreffenden Ökosystems ist [2].

Die (mikrobielle) Biodiversität in natürlichen Ökosystemen kann dabei über genetische Daten abgeschätzt werden. Eine Umweltprobe wird sequenziert, also die DNA-Sequenzen der darin befindlichen Mikroorganismen erfasst und digitalisiert. Auf der Basis dieser Daten können Aussagen darüber getätigt werden, welche Organismen in welcher Anzahl in einem betreffenden Ökosystem vorhanden sind. In den letzten Jahren wurden in mehreren Studien auf der Basis solcher Sequenzdaten sogenannte *co-occurrence networks* berechnet, die als Annäherung an Interaktionsnetzwerke betrachtet werden können. Hierzu wurden verschiedene Methoden eingesetzt, die mit Hilfe von einfachen Korrelationen oder simplen machine learning-Ansätzen numerische Zusammenhänge zwischen einzelnen Organismen im Ökosystem finden [2] und diese auf einen Graph abbilden.

In diesem Projekt sollen möglichst viele der in der Literatur vorhandenen Methoden zur Erzeugung von Interaktions- und *co-occurrence*-Netzwerken in einem R-Paket implementiert werden und die Interaktionen von mehr als zwei Organismen mit Hilfe von *multi-target regressions*-Modellen modelliert werden [3,4]. Dabei sollen zudem typische Fragen aus der Netzwerkmodellierung, wie z.B. Zentralität von Knoten, maximale Subgraphen etc. für diese Problemstellung beantwortet werden.

#### Literatur

[1] Proulx SR et al.: Network thinking in ecology and evolution. Trends in Ecology & Evolution 2005, 20:6.

[2] Karimi B et al.: Microbial diversity and ecological networks as indicators of environmental quality. Environ Chem Lett 2017, 15:265.

[3] Lange A, Jost S, Heider D, Bock C, Budeus B, Schilling E, Strittmatter A, Boenigk J, Hoffmann D: AmpliconDuo: A Split-Sample Filtering Protocol for High-Throughput Amplicon Sequencing of Microbial Communities. PLoS ONE 2015, 10(11): e0141590.

[4] Grossmann L, Jensen M, Heider D, Jost S, Glücksman E, Hartikainen H, Mahamdallie SS, Gardner M, Hoffmann D, Bass D, Boenigk J: Protistan community analysis: key findings of a large-scale molecular sampling. ISME J. 2016, 10(9):2269-79.

Bitte melden Sie sich bei Interesse per Email bei:

Prof. Dr. Dominik Heider - dominik.heider@uni-marburg.de

Stand: 17.04.2018