НЕНАПРАВЛЕННЫЕ МОДЕЛИ И ФАКТОР-ГРАФЫ

Сергей Николенко

НИУ ВШЭ — Санкт-Петербург

08 сентября 2017 г.

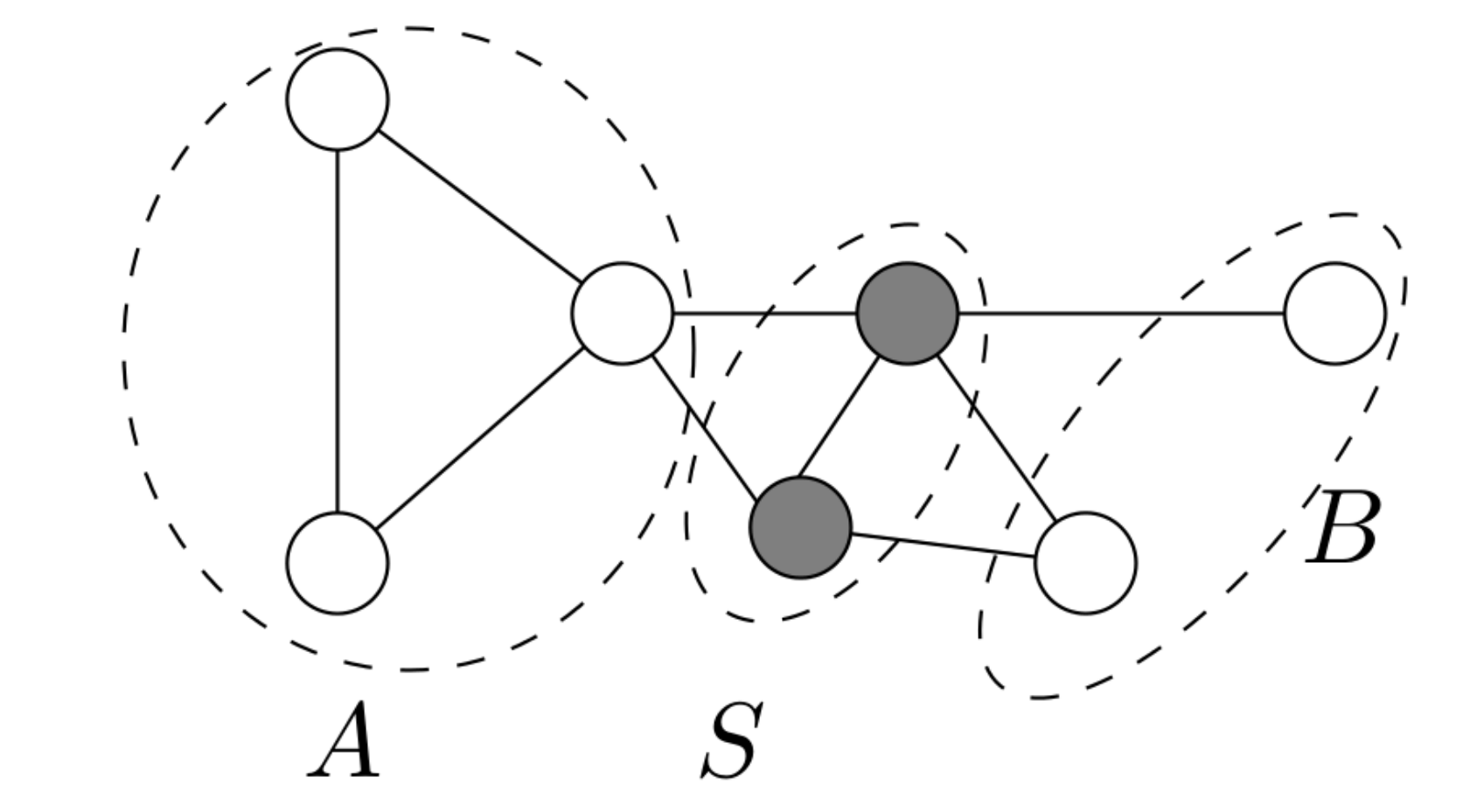
Random facts:

- 8 сентября 1863 г. открылось движение конок по Невскому проспекту
- 8 сентября 1888 г. были сыграны первые матчи в истории английской Футбольной лиги: West Bromwich--Stoke 2:0, Preston--Burnley 5:2, Derby--Bolton 6:3, Everton--Accrington 2:1, Aston Villa--Wolverhampton 1:1

ГРАФИЧЕСКИЕ МОДЕЛИ С НЕНАПРАВЛЕННЫМИ ГРАФАМИ

- Можно сделать и так, чтобы условие независимости было (более) локальным.
- Для этого нужно задавать модели ненаправленными графами. В них условие совсем естественное: множество вершин X условно независимо от множества вершин Y при условии множества вершин Z, если любой путь от X к Y проходит через Z.
- В частности, очевидно, $p(x_i, x_j \mid x_{k \neq i,j}) = p(x_i \mid x_{k \neq i,j}) p(x_j \mid x_{k \neq i,j})$ тогда и только тогда, когда x_i и x_j не соединены ребром.
- Такие модели называются *марковскими сетями* (Markov random fields).

УСЛОВНАЯ НЕЗАВИСИМОСТЬ В НЕНАПРАВЛЕННЫХ МОДЕЛЯХ



ГРАФИЧЕСКИЕ МОДЕЛИ С НЕНАПРАВЛЕННЫМИ ГРАФАМИ

- Поэтому в ненаправленных моделях локальные распределения соответствуют кликам в графе, и факторизация получается в виде

$$p(x_1, \dots, x_k) = \frac{1}{Z} \prod \psi_C(x_C),$$

где C – максимальные клики, ψ_C – неотрицательные функции (*потенциалы*), а Z – нормировочная константа (partition function).

- Поскольку $\psi_C \geq 0$, их обычно представляют как экспоненты:

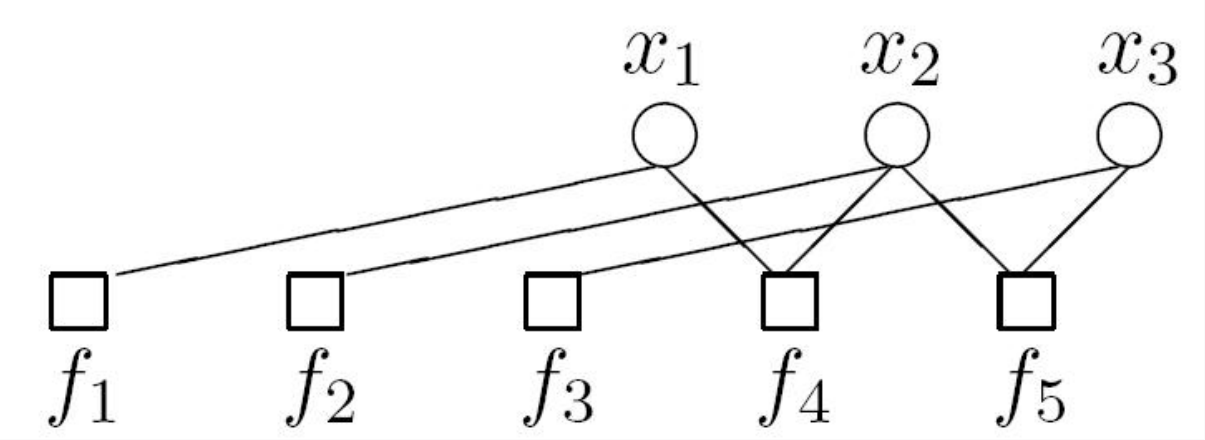
$$\psi_C(x_C) = \exp\left(-E_C(x_C)\right),$$

E_C – функции энергии, они суммируются в полную энергию системы (это всё похоже на статистическую физику, отсюда и терминология).

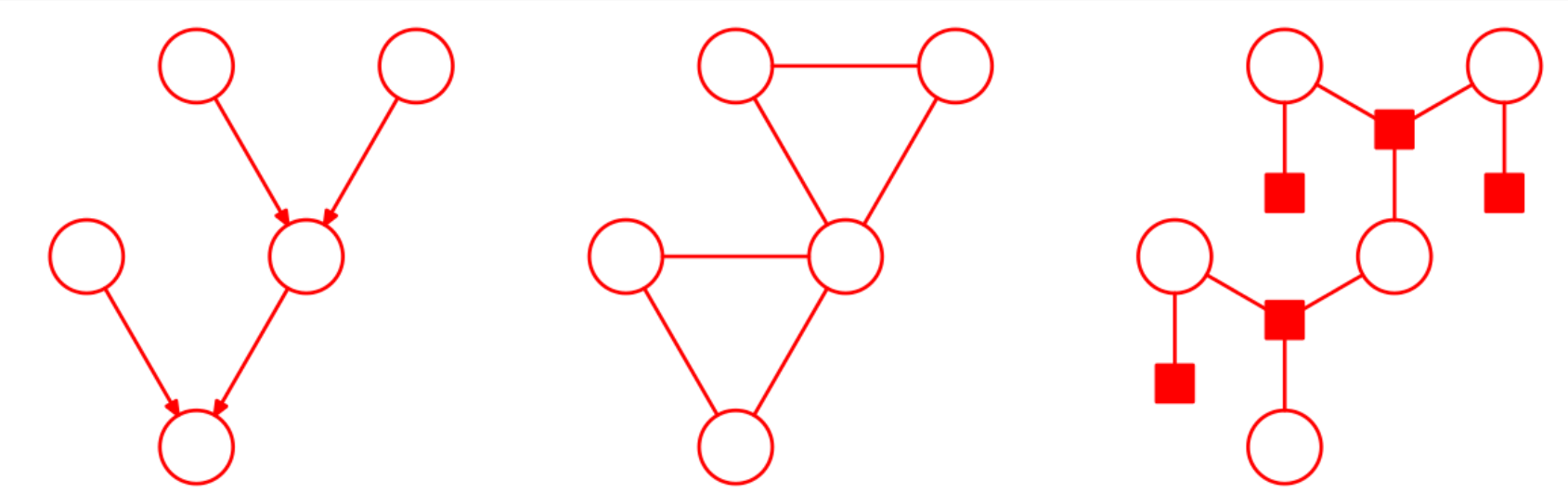
- Интересный факт: назовём *идеальной картой* (perfect map) распределения D графическую модель G, если все условные независимости, присутствующие в D, отображены в G, и наоборот (ничего лишнего). Тогда идеальные карты в виде направленных моделей существуют не у всех распределений, в виде ненаправленных тоже не у всех, и эти множества существенно различаются (бывают распределения, которые нельзя идеально выразить направленной моделью, но можно ненаправленной, и наоборот).

ФАКТОР-ГРАФЫ

- Важная для вывода модификация – *фактор-граф* (можно построить и по направленной модели, и по ненаправленной).
- Фактор-граф – двудольный граф функций и переменных.
- Функция, соответствующая графу, – произведение всех входящих в него функций (т.е. то самое разложение и есть).
- Пример: $p(x_1, x_2, x_3) = f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_1, x_2) f_5(x_2, x_3)$.



ТРИ ПРЕДСТАВЛЕНИЯ



ФУНКЦИЯ В ОБЩЕМ ВИДЕ

- Чтобы поставить задачу в общем виде, рассмотрим функцию

$$p^*(X) = \prod_{j=1}^{m} f_j(X_j),$$

  где $X = \{x_i\}_{i=1}^n$, $X_j \subseteq X$.
- Т.е. мы рассматриваем функцию, которая раскладывается в произведение нескольких других функций.

ЗАДАЧИ

- Задача нормализации: найти $Z = \sum_X \prod_{j=1}^m f_j(X_j)$.
- Задача маргинализации: найти

$$p_i^*(x_i) = \sum_{k \neq i} p^*(X).$$

  Также может понадобиться, например, $p_{i_1 i_2}$, но реже.
- Поиск гипотезы максимального правдоподобия:

$$\mathbf{x}^* = \arg\max_X p(X).$$

ЗАДАЧИ

- Все эти задачи NP-трудные.
- То есть, если мир не рухнет, сложность их решения в худшем случае возрастает экспоненциально.
- Но можно решить некоторые частные случаи.

ПРИМЕР

- Давайте начнём с графа в виде (ненаправленной) цепи:

$$p(x_1, \dots, x_n) = \frac{1}{Z}\psi_{1,2}(x_1, x_2) \dots \psi_{n-1,n}(x_{n-1}, x_n).$$

- Мы хотим найти

$$p(x_k) = \sum_{x_1} \dots \sum_{x_{k-1}} \sum_{x_{k+1}} \dots \sum_{x_n} p(x_1, \dots, x_n).$$

ПРИМЕР

- Очевидно, тут можно много чего упростить; например, справа налево:

$$\sum_{x_n} p(x_1, \dots, x_n) =$$
$$= \frac{1}{Z}\psi_{1,2}(x_1, x_2) \dots \psi_{n-2,n-1}(x_{n-2}, x_{n-1}) \sum_{x_n} \psi_{n-1,n}(x_{n-1}, x_n).$$

- Эту сумму можно вычислить отдельно и продолжать в том же духе справа налево, потом аналогично слева направо.

ПРИМЕР

- В итоге процесс сойдётся на узле x_k, куда придут два «сообщения»: слева

$$\mu_\alpha(x_k) = \sum_{x_{k-1}} \psi_{k-1,k}(x_{k-1}, x_k) \left[\dots \sum_{x_2} \psi_{2,3}(x_2, x_3) \left[\sum_{x_1} \psi_{1,2}(x_1, x_2) \right] \dots \right],$$

справа

$$\mu_\beta(x_k) = \sum_{x_{k+1}} \psi_{k,k+1}(x_k, x_{k+1}) \left[\dots \left[\sum_{x_n} \psi_{n-1,n}(x_{n-1}, x_n) \right] \dots \right].$$

- Каждую частичную сумму можно рассматривать как «сообщение» от узла к своему соседу, причём это сообщение – функция от соседа.

THANK YOU!

Thank you for your attention!