Investigation of Causal Effects

Daniel Lindner, Simon Witzke

6/30/2020

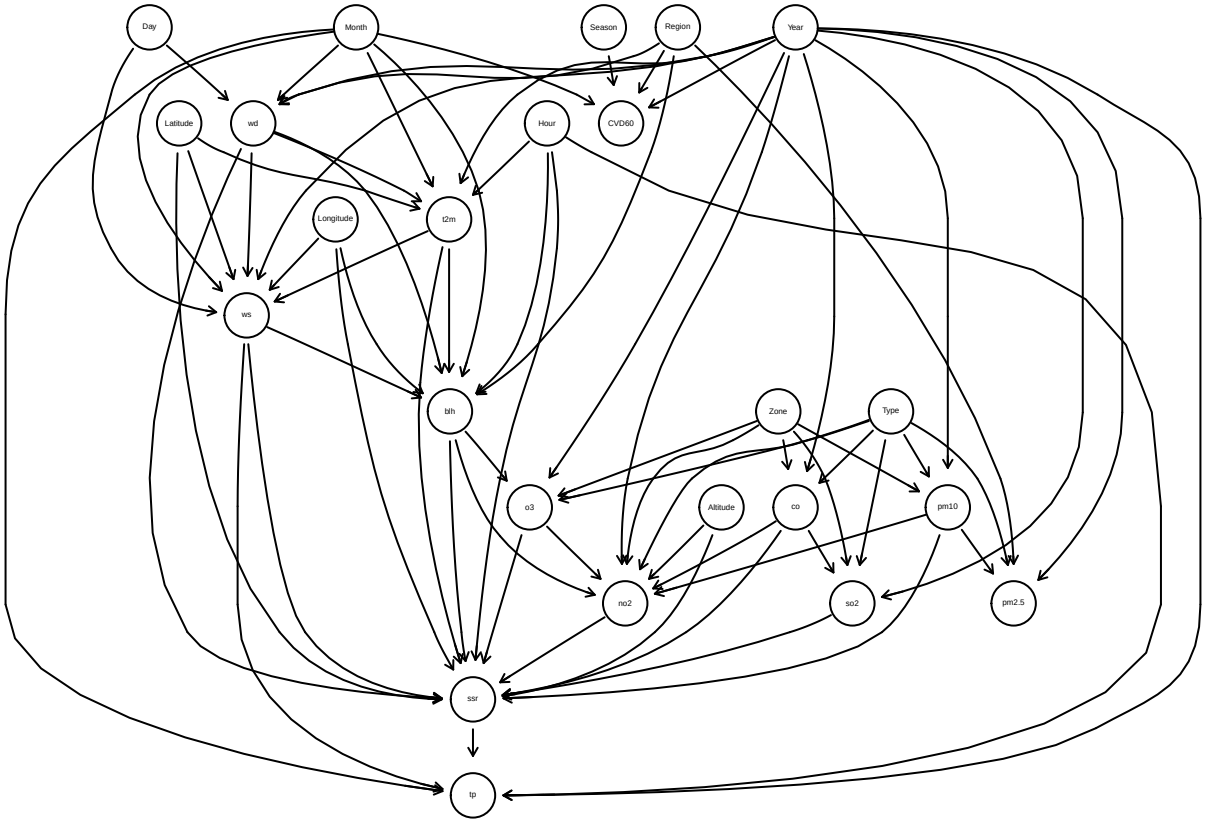
Translation of learned bayesian network to igraph

```
sample.size <- 5e+06
sample <- generate.sample(sample.size)

clustered.data <- sample %>% base::sapply(function(x) if (!is.factor(x)) {
    factor(Ckmedian.1d.dp(x, k = c(3, 5))$cluster)
} else x) %>% base::as.data.frame()

# TODO replace bn with our learned and saved model
adj.matrix.from.bngraph <- amat(bn)
igraph.from.bnmatrix <- graph_from_adjacency_matrix(adj.matrix.from.bngraph)
# TODO mark undirected edges as such
```

Plot network



Find causal effect

```
effect.igraph.from.bnmatrix <- causal.effect(y = "so2", x = "co",
    z = NULL, G = igraph.from.bnmatrix, expr = T)
```

‘$\sum_{Zone,Type,Year} P(so2|Zone,Type,Year,co)P(Year)P(Type)P(Zone)$‘

Estimate causal effect strength

```
causal.effect.probabilites.so2.co <- function(data) {

    data %>% select(so2, co, Type, Zone, DateTime) %>% add_count(Zone) %>%
        mutate(prob_zone = n/nrow(data)) %>% select(-n) %>% add_count(Type) %>%
        mutate(prob_type = n/nrow(data)) %>% select(-n) %>% add_count(DateTime) %>%
        mutate(prob_datetime = n/nrow(data)) %>% select(-n) %>%
        group_by(Zone, Type, DateTime, co) %>% add_count(so2) %>%
        mutate(so2_cond_prob = n/nrow(data)) %>% select(-n) %>%
        mutate(total_prob = prob_zone * prob_type * prob_datetime *
            so2_cond_prob) %>% distinct() %>% group_by(so2, co) %>%
        summarise(`P(so2|Type, Zone, DateTime, co)P(Type)P(Zone)P(DateTime)` = sum(total_prob)) %>%
        inner_join(data %>% group_by(co) %>% count(co) %>% mutate(`P(co)` = n/nrow(data)) %>%
            select(-n))
}
```

```
causal.effect.size.so2.co <- function(data) {
    sum(causal.effect.probabilites.so2.co(data) %>% base::apply(MARGIN = 1,
        function(x) as.numeric(x[3]) * as.numeric(x[4]) * log(as.numeric(x[3])/sum(causal.effect.probabi
            base::apply(MARGIN = 1, function(x) as.numeric(x[3]) *
                as.numeric(x[4])))))
}

xtable(causal.effect.probabilites.so2.co(clustered.data), digits = c(0,
    0, 0, 10, 5))
```

% latex table generated in R 3.6.2 by xtable 1.8-4 package % Wed Jul 01 09:57:24 2020

	so2	co	P(so2|Type, Zone, DateTime, co)P(Type)P(Zone)P(DateTime)	P(co)
1	1	1	0.0000000005	0.07581
2	1	2	0.0000000008	0.10108
3	1	3	0.0000000007	0.10969
4	1	4	0.0000000900	0.67107
5	1	5	0.0000000017	0.04235
6	2	1	0.0006091325	0.07581
7	2	2	0.0006955428	0.10108
8	2	3	0.0005066126	0.10969
9	2	4	0.0035689610	0.67107
10	2	5	0.0004706730	0.04235
11	3	1	0.0000083758	0.07581
12	3	2	0.0000000277	0.10108
13	3	3	0.0000000062	0.10969
14	4	1	0.0000000127	0.07581
15	4	2	0.0000000816	0.10108
16	4	3	0.0000016898	0.10969
17	4	4	0.0000001014	0.67107
18	4	5	0.0000000214	0.04235
19	5	2	0.0000000117	0.10108
20	5	3	0.0000000281	0.10969
21	5	4	0.0000014091	0.67107
22	5	5	0.0000000405	0.04235

```
causal.effect.size.so2.co(clustered.data)
```

[1] 0.0004713883